

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2005, 17:51:02 ; Search time 229 Seconds  
(without alignments)  
3741.085 Million cell updates/sec

Title: US-09-889-874A-23  
Perfect score: 8879  
Sequence: 1 VIKFKLFRRTMSDNEP.....PRKILGRTEKTKPKTRP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot\_03:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	8815	99.3	1660	Q9EVR7	Q9evr7 xenorhabdus
2	1632	18.4	1625	Q7MB38	Q7mb38 photorhabdu
3	1614	18.2	1562	Q883V9	Q883v9 pseudomonas
4	1598	18.0	1385	Q7NWU2	Q7nwu2 chromobacte
5	1583	17.8	1775	Q7N4A7	Q7n4a7 photorhabdu
6	1533	17.5	773	Q9S6J1	Q9s6j1 coxiella bu
7	1530.5	17.2	774	Q52880	Q52880 coxiella bu
8	1416	15.9	1632	Q88LP7	Q88lp7 pseudomonas
9	1214	13.7	709	Q45948	Q45948 coxiella bu
10	1037	11.7	1290	Q88LP2	Q88lp2 pseudomonas
11	932.5	10.5	982	Q88JW6	Q88jw6 pseudomonas
12	920.5	10.4	1669	Q87VG6	Q87vg6 pseudomonas
13	919.5	10.4	505	Q45905	Q45905 coxiella bu
14	919.5	10.4	526	Q9X626	Q9x626 coxiella bu
15	916.5	10.3	526	Q52883	Q52883 coxiella bu
16	912.5	10.3	528	Q45949	Q45949 coxiella bu
17	884	10.0	589	Q45902	Q45902 coxiella bu
18	648	7.3	762	Q93JH8	Q93jh8 vibrio pomm
19	552	6.2	370	Q7NVR8	Q7nvr8 chromobacte
20	486	5.5	389	Q7NVA0	Q7nva0 chromobacte
21	478.5	5.4	310	Q45904	Q45904 coxiella bu
22	450.5	5.1	2217	Q8TP72	Q8tp72 methanosarc
23	450	5.1	380	Q7NWO5	Q7nwo5 chromobacte
24	439	4.9	2554	Q72RQ5	Q72rq5 leprospira
25	432	4.9	2334	1 WAPA_BACSU	Q07833 bacillus su
26	429.5	4.8	336	Q45946	Q45946 coxiella bu
27	420.5	4.7	2364	Q82R58	Q82r58 streptomyce
28	418.5	4.7	1825	Q6MD86	Q6md86 parachlamyd
29	414.5	4.7	2221	Q73BZ3	Q73bz3 bacillus ce
30	396	4.5	1513	Q7NY44	Q7ny44 chromobacte
31	395	4.4	820	Q82QR1	Q82qr1 streptomyce

RESULT 1

Q9EVR7

ID Q9EVR7 PRELIMINARY; PRT; 1660 AA.

AC Q9EVR7; DT 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Nematicidal protein 2.

GN Name=xnp2;

OS Xenorhabdus bovienii.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Xenorhabdus.

OX NCBI\_TaxID=40576;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=I73;

RA Morgan J.A.W., Ousley M., Ellis D., Jarrett P.;

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ296651; CAC19493.1; -

DR InterPro; IPR006530; YD.

DR Pfam; PF05593; RHS repeat; 5.

DR TIGRPFAMS; TIGR01643; YD repeat 2x; 4.

SQ SEQUENCE 1660 AA; 185738 MW; AFABEA20AD70B164 CRC64;

Query Match 99.3%; Score 8815; DB 2; Length 1660;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 MDSNNEFFTQANNFTSAVSGGVDPRGTGLYNIQITLGHIVGNGMLGPTLPLTLSYSPLNKT 73

DB 1 MDSNNEFFTQANNFTSAVSGGVDPRGTGLYNIQITLGHIVGNGMLGPTLPLTLSYSPLNKT 60

QY 74 DIGFGIGFNGFGLSVYDRKNSLLSLSTGENYKVETDKTKLQOKKLDNLRFEKDLKENCY 133

DB 61 DIGFGIGFNGFGLSVYDRKNSLLSLSTGENYKVETDKTKLQOKKLDNLRFEKDLKENCY 120

QY 134 RIHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDWNFEATPRLNRIYDDLDGHD 193

DB 121 RIHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDWNFEATPRLNRIYDDLDGHD 180

QY 194 IPLLNLLEYGGLIKTILTLFPQKQEGYRTLRFLNRLNSIHNFPSLGNENPLTWSPGYTPI 253

DB 181 IPLLNLLEYGGLIKTILTLFPQKQEGYRTLRFLNRLNSIHNFPSLGNENPLTWSPGYTPI 240

QY 254 GKGNGILGQWITSMTAPGGLKETVYNNNGHHPQSANLPVLPYVTLKQVPGAGOPAI 313

DB 241 GKGNGILGQWITSMTAPGGLKETVYNNNGHHPQSANLPVLPYVTLKQVPGAGOPAI 300

QY 314 QAESYTSYHNHYGGSGNGIWNKLDNLYGLMTEYNYGSTESRRYKDKEGHDQIVRIERTY 373

DB 301 QAESYTSYHNHYGGSGNGIWNKLDNLYGLMTEYNYGSTESRRYKDKEGHDQIVRIERTY 360

QY 374 NNHLLTSCCKQONGYIQTETAYYAIIGHNFDSPQSPQLPKTKTETWRSANSVRSEI 433

DB 374 NNHLLTSCCKQONGYIQTETAYYAIIGHNFDSPQSPQLPKTKTETWRSANSVRSEI 433

ALIGNMENTS

Db 361 NNHLLTSECKQNGYIQTETAYAIIGHNFDSPQSQFOLPKTKTETWESADNSYRSEI 420  
Qy 434 TETTFDSGNGPLTKVIKDKTKTKIISPSHTWEYPPAGEVDNCPPEYGTTRFVKKLIQT 493  
Db 421 TETTFDSGNGPLTKVIKDKTKTKIISPSHTWEYPPAGEVDNCPPEYGTTRFVKKLIQT 480  
Qy 494 PYDSEFKDDPEKFIQYRSLIGSOSHVTLLKIEERHYSATOLLNSTLFQYNDKSELGRLL 553  
Db 481 PYDSEFKDDPEKFIQYRSLIGSOSHVTLLKIEERHYSATOLLNSTLFQYNDKSELGRLL 540  
Qy 554 KQTECTKGNGKYVSVVHKTYTKQDDTLQOQSHSIITHDNFTIHRQVRSRYTGRLFSDT 613  
Db 541 KQTECTKGNGKYVSVVHKTYTKQDDTLQOQSHSIITHDNFTIHRQVRSRYTGRLFSDT 600  
Qy 614 DTKDIVTQMSYDKLGRLLTTLNSGTPYANTLYDYELANNLODNNRPPFVIITTDVNGNQ 673  
Db 601 DTKDIVTQMSYDKLGRLLTTLNSGTPYANTLYDYELANNLODNNRPPFVIITTDVNGNQ 660  
Qy 674 LRNEFDGAGRHVSQCLKSDGDKGYTIHTQQYDEQGRHHTSTYSYDLTNGRQOTDPPDKV 733  
Db 661 LRNEFDGAGRHVSQCLKSDGDKGYTIHTQQYDEQGRHHTSTYSYDLTNGRQOTDPPDKV 720  
Qy 734 HLSMSKSYDNWGOIANTHSYGVSEKIVDPITLTKATKQLOSNVNVQKGVVTPSPQ 793  
Db 721 HLSMSKSYDNWGOIANTHSYGVSEKIVDPITLTKATKQLOSNVNVQKGVVTPSPQ 780  
Qy 794 QPIQITLDFEAGHLQSCHTLTRGDWRVRKETDAIGQCTIYOYDYNRVIOITLDPDGTIV 853  
Db 781 QPIQITLDFEAGHLQSCHTLTRGDWRVRKETDAIGQCTIYOYDYNRVIOITLDPDGTIV 840  
Qy 854 NRKYAPFSTDTLITDIRVNGISLGQOTFPLGRLTQSDGGRVWATYSAGNDQCFSTVI 913  
Db 841 NRKYAPFSTDTLITDIRVNGISLGQOTFPLGRLTQSDGGRVWATYSAGNDQCFSTVI 900  
Qy 914 TPDGQFIHQYOPELDDAVLOVANSNEITQFQSNVPTGALLKAVAEQSLTIPIYPSGRL 973  
Db 901 TPDGQFIHQYOPELDDAVLOVANSNEITQFQSNVPTGALLKAVAEQSLTIPIYPSGRL 960  
Qy 974 KMENINDMKMSYLTWLRGLNGYTDLTGTIQKISRDTGHRVTQIKDSSIKITTLNVDDL 1033  
Db 961 KMENINDMKMSYLTWLRGLNGYTDLTGTIQKISRDTGHRVTQIKDSSIKITTLNVDDL 1020  
Qy 1034 RHIGSOVTDLAGHMLTTTVEFDGLNREIGRKLCDSSGHTLDIQOSWLKTOQLANRIVKL 1093  
Db 1021 RHIGSOVTDLAGHMLTTTVEFDGLNREIGRKLCDSSGHTLDIQOSWLKTOQLANRIVKL 1080  
Qy 1094 NGVLQRTQEQSYDSNRNLNOYKCDGAECPDKVGHISIVTONFTYDIYGNITACHTTFADG 1153  
Db 1081 NGVLQRTQEQSYDSNRNLNOYKCDGAECPDKVGHISIVTONFTYDIYGNITACHTTFADG 1140  
Qy 1154 TEDHATFKFANPTDPCQLTEVHTHPDMPDNIRLKYDKAGRVINITDNHGNTEFNFTYDTL 1213  
Db 1141 TEDHATFKFANPTDPCQLTEVHTHPDMPDNIRLKYDKAGRVINITDNHGNTEFNFTYDTL 1200  
Qy 1214 GRLOQGGSVGYDPLNLVLSOKTDTLCELYYRETMLVNEVRNGEMIRLLRTGTTIAQ 1273  
Db 1201 GRLOQGGSVGYDPLNLVLSOKTDTLCELYYRETMLVNEVRNGEMIRLLRTGTTIAQ 1260  
Qy 1274 QRASKVLLTGTDSQOSVILITSDKQNSQAEYSAYGKHKSTANDASILGYNGERADPVSGV 1333  
Db 1261 QRASKVLLTGTDSQOSVILITSDKQNSQAEYSAYGKHKSTANDASILGYNGERADPVSGV 1320  
Qy 1334 THLNGYRSYDPTLMRFTPPDSLSPFGAGGINPYSCYCLGDPINRSDPSGHLISQAWTIG 1393  
Db 1321 THLNGYRSYDPTLMRFTPPDSLSPFGAGGINPYSCYCLGDPINRSDPSGHLISQAWTIG 1380  
Qy 1394 MGIAGLLLTIATGMAIAAGGIAAIASTSTTALAFGALSVTSDITSIIVSGALEDASPK 1453  
Db 1381 MGIAGLLLTIATGMAIAAGGIAAIASTSTTALAFGALSVTSDITSIIVSGALEDASPK 1440  
Qy 1454 ASSILGWVSMGMGAAGLAESAIKGOTKLATHLGAFAEDGENALLKSTSESSIRIKGWVTRS 1513  
Db 1441 ASSILGWVSMGMGAAGLAESAIKGOTKLATHLGAFAEDGENALLKSTSESSIRIKGWVTRS 1500

Qy 1514 LDRBIVRNEEGQVIKDHSGRYTDNFMGKGEQAILVHGDKDGFLYHTEGKNHNGKGPYTRH 1573  
Db 1501 LDRBIVRNEEGQVIKDHSGRYTDNFMGKGEQAILVHGDKDGFLYHTEGKNHNGKGPYTRH 1560  
Qy 1574 TPEQLVDYLKONNIVDLTQGGDKPVLHLLSCYKSSGAADKMAKYNRPVIAVSNKPTISQ 1633  
Db 1561 TPEQLVDYLKONNIVDLTQGGDKPVLHLLSCYKSSGAADKMAKYNRPVIAVSNKPTISQ 1620  
Qy 1634 GLARIERKDFLLKSTYHSYDPRKIIILGRTEKTVKPKTRP 1673  
Db 1621 GLARIERKDFLLKSTYHSYDPRKIIILGRTEKTVKPKTRP 1660  
RESULT 2  
Q7MB38  
ID Q7MB38 PRELIMINARY; PRT; 1625 AA.  
AC Q7MB38;  
DT 01-MAR-2004 (TREMBLrel. 26, Created)  
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Similar to the nematocidal protein 2. Probable membrane protein.  
GN OrderedLocusNames=plu2222;  
OS Photorhabdus luminescens (subsp. laumondii).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_TaxID=141679;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TT01;  
RX MEDLINE=22957627; PubMed=14528314;  
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,  
RA Taurit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,  
RA Dassa E., Derose R., Deruelle S., Freyssinet G., Gaudriault S.,  
RA Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V.,  
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.,  
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus  
RT luminescens".  
RT Nat. Biotechnol. 21:1307-1313(2003).  
RL Nat. Biotechnol. 21:1307-1313(2003).  
DR EMBL; BX571866; CAEL4515.1; -.  
DR Photolista; plu2222; -.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; Rhs repeat; 7.  
DR TIGRFAMs; TIGR01643; YD\_repeat\_2x; 3.  
KW Complete proteome.  
SQ SEQUENCE 1625 AA; 180118 MW; BFA362F4ECB3EA97 CRC64;  
Query Match 18.4%; Score 1632; DB 2; Length 1625;  
Best Local Similarity 31.0%; Pred. No. 1.1e-77;  
Matches 494; Conservative 245; Mismatches 653; Indels 204; Gaps 59;  
Qy 18 NEFTQANNETSAVSGGVDPRTGLYNTQITLGHVGN--GNLGLTLELTLSYSPLNKTDI 75  
Db 4 NDIYSNAPNFGSYINTGVDPRTQGYSANINIILRPNNVGNIEQV--LNLFSPLTLNN 61  
Qy 76 GFGEFGNFGLSVYDRKNLSLSLSTGENYK--VETDKTVKLOKKLDNLRFEKDLKENC 132  
Db 62 GFGEFGNFGLSVYDRKNLSLSLSTGENYK--VETDKTVKLOKKLDNLRFEKDLKENC 120  
Qy 133 YRIHKGSGDLEVLTFGNNAFADLKVPKLLNPNAGHAIYIDWNFEATQPLNRIYDLDLGH 192  
Db 121 FYVYKNGIETLKRIGSS--DIAKTVALEFPDG-----EVFDLIYS 161  
Qy 193 DIPLLNLEYGLIKTILTL--FPQKEGYRTELAPL--NRQLNSIHNSLGNENPLTWSFGY 250  
Db 162 RFAISEIKYRWTKYIKLNYSGNN---CTSVSYPDNNISAKIAFYDRNDYLTITVTPY 218  
Qy 251 TPICK-----NGILQWITSMTAPGLKETVNSNNQCHHPPOSANLPVLP 297  
Db 219 DAGSPIDSARFKMTYQTLKGF--PVISAPRTPTGYVELVSYKEN--GHKVYDTESIPYAA 275  
Qy 298 YVTLMKQVFCAGOPAQAEYSYTS--HNYVGGSGNGI--WANKLDNLYGLMTEYNYGSTESR 355  
Db 298 YVTLMKQVFCAGOPAQAEYSYTS--HNYVGGSGNGI--WANKLDNLYGLMTEYNYGSTESR 355

Db 276 ALTIQ---PENGOPALSKYEVSSVNFGLYSGSRTSFDSSQDNLYLVTKGTYTSIE--- 330  
Qy 356 RYKDEGHDOIVRIERTYNNYHLLTSECKQNGYIQTETAYYAIIGHNFDSPQSQFOLP 415  
Db 331 --RVLGQNVISVTERVDFKFLMTKEAKTQDNKRITTEITYNEDPSKFSQPENLQOP 388  
Qy 416 K-----TKTETWSADNSYRSBITETPTDESNGPLTKVKKTKQKIISPSTHWEYYP 469  
Db 389 SHLVTRYDLOQ-----NTSRESVNIKSDDWGNTLL-ITETSGIQK-----EYVYVPV 436  
Qy 470 AGEVDCPPPEYGFTRFVKKIQTQPYDSEFKDDPEKFIQYRYSVLGSSQSHVTLKIEERH 529  
Db 437 NGEGNPCADPLOGFSLFKSLVTKQSPDAAQSVANRVSITTKQKFTTGAYVK--EYVS 494  
Qy 530 SATOLLNSTL---FOY---NTDKSELGRLLKOTECTKGNGKTSYVVKHFTYTKDDTLQ 583  
Db 495 KASSETIDSKIVRTFNYSPTNKSH-GSLAKITSVMNQ-----QTVTFKYEYDSEMT 548  
Qy 584 QHSITTHDNFIHRSQVRSRTGRLFSDTDKDIVTQMSYDKLGRLLTRTLNSGTPYAN 643  
Db 549 TNSTVTGPDGTMSKKNVTSIYTHRQLRKVDVNVHVIDQSYDLSGRILGIIQIIPGTKEI 608  
Qy 644 TLTYDELNNLDDNRPPVITTDVNGNLNNEFDGGRHVSCCLKQSDG-----D 695  
Db 609 KRSYIYQYPCGDEDFWP-VMEIDSQGIIRKTHYDGMGRICISIEQDDDGWVGTSGIYQ 667  
Qy 696 GFYTIHTQYDEQGRHHTSYSDYLTN-----GROQTDPDVKVHLSMSKSYDNWQIANT 750  
Db 668 GYRKVLARQYDVLGOLVKEISNDLWLDLSANPLFELTTP--LVTKTYQYDGNRVS 725  
Qy 751 HWSYGVSEKIVDPITLTKATK-----QLQSNNNVQTKGVTTVTPSQOPIQITLFDEA 804  
Db 726 EYSDGRILEIHDPTTRITTOGVKGLMLNIQNNP-----EQPASIKVWYPD 773  
Qy 805 CHLOSCHLTRGDWRVREKETAIGCTIYQYDNNRVLIQITLPGTIVNRKYAPSTDT 864  
Db 774 GAIYSTRTRYDGFGRVTTETDAEGYATQIEYDFDRIVKKTLPTRTILESAYAFSHEE 833  
Qy 865 LITDRVNGISLQQTDFGLSRLTOSQDGRVWAVYTSAGNDQCSTVITP--DCQFIHY 922  
Db 834 LISALNVNGTQLGSLVDGLGRVTRTVGGRKTEYLYGSGDK-PIQSVTPAHNKQNDY 892  
Qy 923 QYQPELDDAVLQVANSNEITQFSYNPVTGALLKAVAEGOSLTP1-YYPSSGRKLMENIN-D 980  
Db 893 LY--ALGSVMSKFTTETSQONFSYQKTGALLSA-TEGVSQSNYSYFSGVLQHSFSRD 949  
Qy 981 MKKMS---LWTLRLGLENGYDTLGTI QKISRDRHGRVTOIKDSIKITLNDLNRHIG 1037  
Db 950 NKPISSGDYRYTMSGLIQRHKDSFAHDHVYSYDAEGRVKTBSQSYATFEYDNGRLIT 1009  
Qy 1038 SQVTDLATGHMLTTTVEPDLNREIGRKLCDSSGHTLDIQOSWLKTQOLANRIVKLVYL 1097  
Db 1010 TTTKTTLSQLATKIEYDVPREIKRSLISDFSQV-ITLSYTONQISQIRITSIDGVV 1068  
Qy 1098 QTEQYYSYDSRNRLNQKCDGAECPCTKYGHSIVTQNFYDIYGNITACHTTTFADGTBDH 1157  
Db 1069 MKNERYOYDSNQLSQYCEGQSPVDHTGRVLSQI1YHYDQWGNIKRLDNTYRDKET- 1127  
Qy 1158 ATFKANPTDPCOLTEVHTHDPMDPNILKYDKAGRVINIT-DNHGNTENFTYTLGR 1216  
Db 1128 VDYHFSQ-ADPTQLIRI-----TSDKQOIELSANG---NLTRDBKQGT--LIYQNNRL 1177  
Qy 1217 ---ONGQGSV---YGYDPLNRLVSQ--KDTDLDCELYRETMVNEVRNGEMIRLL--- 1264  
Db 1178 VQVQSKGMLVCOYQYDALNKLTAQVLANGTVNRQ--YYASGNVANVQLGDEITWLSDK 1236  
Qy 1265 -----RTGETIIAQQRASKVLLTGTDSQOSVILTSDKQNLSEAYSAYGKHKSTAND 1316  
Db 1237 QRLGHQSTKNGSVYQY-----GTDHNSVIVASQENELMALSYTPYGFRLI--- 1285  
Qy 1317 ASILGVNGERADPVSGVTHLNGYRSYDPTLMRFTTPDLSLSPFAGGNNPYCYLGDPI 1376  
Db 1286 SSLPLGLNGAQVDVPTGWYFLNGYRVFNPLMRFHSPDSFSGRGGVNPYTYCQGDPI 1345

## RESULT 3

Q883V9  
ID Q883V9 PRELIMINARY; PRT; 1562 AA.  
AC Q883V9;  
DT 01-JUN-2003 (TRENBLrel. 24, Created)  
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE YD repeat protein.  
GN OrderedLocusNames=PSPTO2239;  
OS Pseudomonas syringae (pv. tomato).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=323;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DC3000;  
RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;  
RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,  
RA Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,  
RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,  
RA Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q.,  
RA Khouri H.M., Pedorova N.B., Tran B., Russell D., Berry K.J.,  
RA Uterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,  
RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,  
RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,  
RA Bender C.L., White O., Fraser C.M., Collmer A.  
RT "The complete genome sequence of the Arabidopsis and tomato pathogen  
Pseudomonas syringae pv. tomato DC3000."  
RT Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).  
DR EMBL; AE016863; AA055755.1; -.  
DR TIGR; PSPTO2239; -.  
DR InterPro; IPR000977; DNA\_ligase.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; RHS\_repeat; 6.  
DR TIGRFAMs; TIGR01643; YD repeat 2x; 3.  
DR PROSITE; PS00697; DNA\_LIGASE\_A1; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 1562 AA; 175713 MW; 8DC10DA1BF37BF1 CRC64;

Query Match 18.2%; Score 1614; DB 2; Length 1562;  
Best Local Similarity 30.3%; Pred. No. 9.3e-77;  
Matches 490; Conservative 239; Mismatches 700; Indels 190; Gaps 51;

Qy 14 MSDNNEFTQANNFTSAVSGGVDPRTGLYNTIQTILGHVGNGLGPTLPLTISYPLNKT 73  
Db 1 MTTSTSVHSAFNMSYLSQSGVDPRGTGYTVSINLPEVKSNGLRGPVPLVLSYNPLNVQ 60  
Qy 74 DIGFICIGNFGSLVVDYDRKNSLLSISTGENYKVI-ETDKTVKLQKKLDRPEKDKENC 132  
Db 61 DSGFGLGNWLQSLQYDPCGTRIVSLGSGETPKVSGSLGDLWPEKKLDSFHYKQ-DDTR 119  
Qy 133 YRIIHKSGDIEVLGTGFNNNAFDLKVPKKLANPAGHAIYIDWNFEATQRLNRIYDDLGH 192  
Db 120 YRVVHKSQVBELEVLGSLGNRIALPVRIYSPGEGHITLHVASFAGYQMLSEVDD-DQ 178  
Qy 193 DIPLNLLEYQGLIKTILTLFPQKEGYETELRNLQNLNSIHNFSLGNENPLTWSGFYTP 252  
Db 179 VILITRD----STSVRLLYLGPADKADAEFVMIILSGSNRNVARIELPTANKASWRFTYSI 234







```

Db      137 LYKTSREKLQOL--DRANPAVAVSEIYAPSGHKLCKVKNFVNVHNYNYQLMEVCDAE- 194
QY      192 HDIPLNLEVQGLIKITLTPQKEGYRTLPFLANQLNSHNFSLGNENPLTWSGFGYT 251
Db      195 ---TLKADLATEKIEFTVPWSPSPSYVTLNMTNDLLQTVISAS-----ELTWHLLEYE 246
QY      252 PIKNGILGOWITSMAPGGLKETVANSNNQGHFF----- 288
Db      247 TEGAHKNI---LTKVTPSLIEKVY--HETGHTLTPKCYYPNANGPGLIRODPKS 301
QY      289 -QANILPVLPIYVTLMKQVPGAGAPAOAEYSYTSYSHNYVGGSGNGIWN-----NKLNDLYGL 343
Db      302 STTATVHEFPYVTHHIIAGSGSPQVIRYVSPENFLGKNKNMKDPPLPQQDNAYLA 361
QY      344 MTEVYNGSTESRYKDEKQHDQIVRIERTYNNVHLITSECK-----QONGHIQIOTETAY 397
Db      362 NSYKKTSTEVREYNEKR-----YCHREYNKPHLLVSETVETVTPSPQKLKETIKY 416
QY      398 YAIIGHNFD--SOPSOQLPKTKTETWRSADNS-----YRSEITETTFDESNGNPLTKVIK 450
Db      417 YADVGRSFDNNKPPQFLMPNIVETIWHNPRESSASTTORKEKETTQWEYNAQCNLISWILP 476
QY      451 DKKTQKIISPSHWEYYPAGEV---DNCPPBPYGFTRPVKKI-IOTPYD-SEFKDDPEK 505
Db      477 DMTTKT-----TYIAPDGBETTDHCPABPNGFEREIKETIAVEAPSPPLTKTITILR 529
QY      506 FIOYRSLIGSOS-----HVTLKERHYS-----ATOL-----LNSTLQYNTD 545
Db      530 KVTYNYKSDYVTPPKNQASVVKSMVLPSLTHYSRRDCCADHLEKVKINTVSYENTQ 589
QY      546 KSELGLRLKQETCKTGENG-KTYSVVHKFTYTKQDDTLQOQSHITTHONF-----TIHRS 599
Db      590 NAFLNGRVEORNSYLFENGQNTSYEDYSWENKNGASCICKTKTSGKCGTTPVPSHE 649
QY      600 QVRSYTGRLFDSDTKDIVTQMSYDKLGRLLTRLNGSTPYANTLTVDYELNNLODNR 659
Db      650 QWYSRSTGLIFPKDAQDNNTVFQDTHGLISSTINADTAYEKTVKYAWSYKSR----- 705
QY      660 PPFVITTTDVNGQLNRNEFDGAGRVSOCLKSDGSGFYTHITQOYDEQGRHHTSYSD 719
Db      706 --VTVQTDIHENRYITENDGLRGLPKKSYSPAGQGFDMERYQYNPLGOLQOALISCD 763
QY      720 YLTNGRQOTDPDKVHLSKSYDNNQGOIANTHWSYGVSEKITVDPITLTATKQLQSNMN 779
Db      764 HML-FENKSEKCSMTVSLDYDDGHSRLRYSDGTSVKNTFDPKMTSEHQVSNDES 822
QY      780 VQTKEVTTYPSQOPIQITLFDGAGHLOSCHTLTRDGDWRVREKTDIAQCTIYQYDNY 839
Db      823 QSSGIIRTYNQFGQITATERLTTSRIQGGCWHYLRDELGR-LVSNANGNTLLIAYDAF 881
QY      840 NRVIQITLFDGTIVNRKYA-PESTDLTITDIRVN--GISLGQOTFDGLSRLTQODGGRV 896
Db      882 DRVIKQTFADGTTISWAVENGVSVMASATPGINGQTPVILGTQILDGLGRVIDMESGRK 941
QY      897 WAYTVSAGNQCQSPVITPDQO-----FHYQYQPELDLDAVLQVAS--NEITQQFSYN 947
Db      942 IKLDYE-GASVPVDTVTVYKQPTSGDKDILVHYEPEKLNNAVTKITAGQDDVDQTFKYD 1000
QY      948 PVTGALLKAVAE-----QSLTPYIYPSGRKME-----NINDMKMSYIWLTLR-G 992
Db      1001 PKTGLLTAEIYKKNDEMISSKLIPDTYLSGQLASEKLVSYKNKNRAYMHFTCYFS 1060
QY      993 LENGVTDLTG-----IOKISRDTHGRVQIQIDSSIKTTLNLYDNLNRHIGSQVTL-- 1043
Db      1061 FANRPTAIKSTINGPRYIINKSYDQNGRLTVERNDIEINLKYDKLSRHVCQSTYELFD 1120
QY      1044 ATGHEMLTTFVDFGLNREIGRK-----LCSSGHT-LDIOQSWLTKQQLANRIVKLVNGV 1096
Db      1121 TINNKITTLDDFGRETKETIRRHNGKKNKGITDIHISQTFNEQDKITNKL-LHGT 1179
QY      1097 LQ-RTEQYSYDSRNLNOKYKDGAECPDKYGHISVITQ--NFTYDIGNITACHTTADG 1153

```

```

Db      1180 KOISKEVYTHKRGGLETTYTL-----MELVGEREITQCSYKYDHLGNITQ-HSITTEG 1232
QY      1154 TEDHATFKFANPT--DPCQLTEVHH--THPDMFDIRLYKYDAGRIVITD-NHGNTEPNT 1209
Db      1233 KTTISTYTYGNGNIQDPCQLIDVSTCTGNNSSSLRFTYNGRGALVCENDENNTKIRWT 1292
QY      1210 YDTLGLQLONGQGSV-----YGYDPLNRLV--SOKTDT-LDCELYYRETMVN----- 1253
Db      1293 YDSLGRDITVDALHKVETRYLFDATNRLIIRKSEKNGTPYHDLISYCSNLSLVDNRYFYG 1352
QY      1254 EVRNGEMLR-----LLRTGETITIAO--QRASKVLLTGTDSQSVILSDKONLSQE 1302
Db      1353 EKEDNAADRKYNKVGGICLGFSPQCHQTPTSVASRYTETATDGKSVIAITFQGEDVOHI 1412
QY      1303 AYSAYGKHKSSTANDASILG-----YNGERADPVSGVTHLGNYSRSDPTLM 1348
Db      1413 AYSPWG--VTTQAMVTAGQQPHNTAIEBPFRNGEQWDTASASYLLGNGYRAYRPDLM 1469
QY      1349 RFHTPDSLSRPGAGGINPYSCLDGPINRSDPSGHLSHQAWTGTGIMGIAGLLLIATCGM 1408
Db      1470 RFTAPDSWSPGAGGINAYAYCGGDPVNLNDPSGHSIGWGANITGGIGLLLPFTYGG 1529
QY      1409 AIAAAGGIAAAIASTSTTALAFGALSVTSDITSVSGALEDASPKASSILGWSMGMAA 1468
Db      1530 SLEGLGVAARGLT-----ALDAASGVTATASGALENKNPETSRRLGWMISLGLG-- 1579
QY      1469 GLASAIKGGTKLA-----THLGAF-----AEDGENALLKSTSE-- 1502
Db      1580 --LFSMWIGGYSLAQWVNVRLTNSFRTPYHYPTSLGEVNLRSKSSDWVNNARRSLNSGENW 1637
QY      1503 -----SSRIKWG--VTRSLDREIVNEBGOVIKDHSRGVTONFMGKGEQAI-----L 1547
Db      1638 HSEVLGNRIWNGDTKIRGLD-----IKYPLEQISRRPSNG--DIVLLSGSHGVQNGDNWL 1692
QY      1548 VHGDQDGLYH-----TEGNKHNKGPTVTRHTPEQLVLDVILKNNIVDLTQGGDKPV 1598
Db      1693 INGSRRGSLHLHFPFKSDMTVYGGSWKGRVTVHRN-----LATMSEIDFGTLLNNGNSHI 1747
QY      1599 HLLSCYKSSGA 1610
Db      1748 ILGYCYGRNDQA 1759

```

## RESULT 6

```

Q9S6J1 ID Q9S6J1 PRELIMINARY; PRT; 773 AA.
AC Q9S6J1 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Coxiella burnetii.
OG Plasmid QpDV.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI140;
RA Radomski K.U., Willems H., Lautenschlaeger S., Jaeger C., Baljer G.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF131076; AAD33495.1; --
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; Rhs repeat; 6.
DR TIGRfams; TIGR01643; YD repeat_2x; 2.
KW Hypothetical protein; Plasmid.
SQ SSQUENCE 773 AA; 83759 MW; 4CSFCE0481CCAC44 CRC64;

```

Query Match 17.5%; Score 1553; DB 2; Length 773;  
 Best Local Similarity 44.6%; Pred.No. 5.5e-74;  
 Matches 342; Conservative 119; Mismatches 275; Indels 30; Gaps 17;

QY 741 YDNWGOIANTHWSYGVSEKITVDPITLTATKQLQSNNNVQTGKEVTTYTPSQOPIQITL 800

Db 4 YDSWGQNLHVTFSVSDGQERSVYDPIITRRAT--LQPSGSGKLQQLTEYNLAGLPIKVTQ 61  
QY 801 FDEAGHLQSCHTLTRDGDWRVRKETAIGCTIYQYDNNRVQITLPDGTIVNRKYAPF 860  
Db 62 YDSQGTQEGSAHYEDGLQKRETDDELQVTLVEYDFHGRVQTTLTPENTIIQKSYAPH 121  
QY 861 STDTLTDIRVNGISLQGOQTFDGLSRLTOSQDGRVWAYTYSAGNDCQCPSTVITPDGQFI 920  
Db 122 STASLTIGISVNNFSGNQTDFSLERLTETTSQRTSAFSEYENAS-SVPAAVTAPTGETV 180  
QY 921 HYQYQPELDDAVLQVANSNEITQFSSYNPVTGALLKAV-AEQSLTIPIYPSGLKWE-NI 978  
Db 181 SYELKELGNVAKKISAPEILKTWDYDALTGAMTSATQAAGMIQMTYTPSGLLKNETSM 240  
QY 979 ND---MKQMSYLTWLRGLNGYTDLTGTQIKISRDTHGRVTQIKDSSIKTTLNYDDLNRH 1035  
Db 241 PDGAQKSTAYTYSLAGAPQSYTDVFGVTRYDYDEHGRIGIEDNDIKVSLGYDAFGRF 300  
QY 1036 ISQVTDLATGMLTITTVBFDGLNREIGRKLCDSSGHTLIDIQSWLKTQQLANRIVKJNG 1095  
Db 301 TKQATDKKTGAVLSTLTLYDDLNRKREISASQSVLVIEQYQRNHLLKERITQGR 360  
QY 1096 VLQRTQOYSVDSNRNLNOYKDCGACPTDKYGHISIVTQNTFYDNIYGNITACHTTTADGTE 1155  
Db 361 TTLRKEVFAVDSNRNLIEYTCNGEARPODPYGKAIHRTQFSYDALGNMTKTQTFSGG-R 419  
QY 1156 DHATPKFANPTDPCQLTEVHHTHPMDNIRLKYDKAGRVINITDNGHNTENFTYDTLGR 1215  
Db 420 NTATYIY-SAIDPTQLKVNNDHSDYKPEITLEYDKAGRMIR--DEAGRT--LRYDALG 474  
QY 1216 LQ--NG---QGSVYGYDPLNRLVSQKT-DTLDCELYYRETMLVNEVR--NGEMIRLLRTG 1267  
Db 475 LQVNGAGAKGGQYAYDALNTLVSVQVQDEPIYDLYRADDLVGEARRDSSQTRVYKSN 534  
QY 1268 ETIIAQ---ORASKV--LLTGTDSQSVIILTSQKNSQEA-YSAVGHKSTANDASILGY 1322  
Db 535 GCCVGCTQKGSNTSRLTTTQOGSVLSVSEGNHAPQDCIYTRYGRTPTETPSVLGF 594  
QY 1323 NGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFGAGGINPYCYCLGDPINRSDPSG 1382  
Db 595 NGERLDPVSGVTHLNGYRAYNPILMRFNCPDSWSPFGAGGINPYAYCDGDPINRVDPN 654  
QY 1383 HLSQAWTIGIGMGIAGLLTITATGMAIAAAGGIAAIASTSTTALAFGALSVDTSITSI 1442  
Db 655 HLSWQAEGLGLGVGLVLAFTAGTSIAAAGASIAESASISLVVGTGLGAADVASI 714  
QY 1443 VSGALEDASPKASSILGWVSMGMAAGLAE---SAIKGGTKLATHL 1485  
Db 715 ASGALEDANPOASATLGIWISLGLGPGAVSGLATARAAGKLLISGL 760

## RESULT 7

OS2880 PRELIMINARY; PRT; 774 AA.  
ID AC O52880;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein orf 774.  
GN Name=orf 774;  
OS Coxiella burnetii.  
OG Plasmid QpRS.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;  
OC Coxiellaceae; Coxiella.  
OX NCBI\_TaxID=777;  
[1]  
RP SEQUENCE FROM N.A.  
RA Lautenschlaeger S., Jaeger C., Willems H., Baljer G.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y15898; CAA75841.1; -.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; RHS\_repeat; 6.

DR TIGRFAM; TIGR01643; YD\_repeat\_2x; 2.  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 774 AA; 83774 MW; 918A54333D7991BD8 CRC64;

Query Match 17.2%; Score 1530.5; DB 2; Length 774;  
Best Local Similarity 44.5%; Pred. No. 8.6e-73;  
Matches 341; Conservative 120; Mismatches 275; Indels 31; Gaps 18;

QY 741 YDNWGQIANHWSYGVSEKITVDPIITLTATKOLQSNNNVOTGKEVTTTPSQOPIQITL 800  
Db 4 YDSWGQNLHVTFSVSDGQERSVYDPIITRRAT--LQPSGSGKLQQLTEYNLAGLPIKVTQ 61  
QY 801 FDEA-GHLQSCHTLTRDGDWRVRKETAIGCTIYQYDNNRVQITLPDGTIVNRKYAP 859  
Db 62 YDSQGTQEGSAHYEDGLQKRETDDELQVTLVEYDFHGRVQTTLTPENTIIQKSYAP 121  
QY 860 FSTDTLTDIRVNGISLQGOQTFDGLSRLTOSQDGRVWAYTYSAGNDCQCPSTVITPDGQF 919  
Db 122 STASLTIGISVNNFSGNQTDFSLERLTETTSQRTSAFSEYENAS-SVPAAVTAPTGET 180  
QY 920 IHVQYQPELDDAVLQVANSNEITQFSSYNPVTGALLKAV-AEQSLTIPIYPSGLKWE-N 977  
Db 181 SYELKELGNVAKKISAPEILKTWDYDALTGAMTSATQAAGMIQMTYTPSGLLKNETS 240  
QY 978 IND---MKQMSYLTWLRGLNGYTDLTGTQIKISRDTHGRVTQIKDSSIKTTLNYDDLNR 1034  
Db 241 MPDGAQKSTAYTYSLAGAPQSYTDVFGVTRYDYDEHGRIGIEDNDIKVSLGYDAFGR 300  
QY 1035 HTGSQVTDLATGMLTITTVBFDGLNREIGRKLCDSSGHTLIDIQSWLKTQQLANRIVKLN 1094  
Db 301 TKQATDKKTGAVLSTLTLYDDLNRKREISASQSVLVIEQYQRNHLLKERITQGR 360  
QY 1095 VLQRTQOYSVDSNRNLNOYKDCGACPTDKYGHISIVTQNTFYDNIYGNITACHTTTADGT 1154  
Db 361 TTLRKEVFAVDSNRNLIEYTCNGEARPODPYGKAIHRTQFSYDALGNMTKTQTFSGG- 419  
QY 1155 DHATPKFANPTDPCQLTEVHHTHPMDNIRLKYDKAGRVINITDNGHNTENFTYDTL 1214  
Db 420 NTATYIY-SAIDPTQLKVNNDHSDYKPEITLEYDKAGRMIR--DEAGRT--LRYDALG 474  
QY 1215 LQ--NG---QGSVYGYDPLNRLVSQKT-DTLDCELYYRETMLVNEVR--NGEMIRLLRT 1266  
Db 475 LQVNGAGAKGGQYAYDALNTLVSVQVQDEPIYDLYRADDLVGEARRDSSQTRVYKS 534  
QY 1267 ETIIAQ---ORASKV--LLTGTDSQSVIILTSQKNSQEA-YSAVGHKSTANDASILG 1321  
Db 535 GCCVGCTQKGSNTSRLTTTQOGSVLSVSEGNHAPQDCIYTRYGRTPTETPSVLG 594  
QY 1322 NGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFGAGGINPYCYCLGDPINRSDPS 1381  
Db 595 NGERLDPVSGVTHLNGYRAYNPILMRFNCPDSWSPFGAGGINPYAYCDGDPINRVDPN 654  
QY 1382 HLSQAWTIGIGMGIAGLLTITATGMAIAAAGGIAAIASTSTTALAFGALSVDTSITSI 1441  
Db 655 HLSWQAEGLGLGVGLVLAFTAGTSIAAAGASIAESASISLVVGTGLGAADVASI 714  
QY 1442 VSGALEDASPKASSILGWVSMGMAAGLAE---SAIKGGTKLATHL 1485  
Db 715 ASGALEDANPOASATLGIWISLGLGPGAVSGLATARAAGKLLISGL 761

## RESULT 8

OS2880 PRELIMINARY; PRT; 1632 AA.  
ID AC Q88LP7;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocusNames=PP1882;  
OS Pseudomonas putida (strain KT2440).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.









1272 AQO---RASKVLTLTGDSQSVLTSDKQNLQSOAYSAVYKHKSTANDASILGYNGERAD 1328  
 617 GQOIVDDAEQTLILLTLDANQVMEFQOQQLKAVYSAYGERHSEBALLSTAGFNGEVRE 676  
 1329 PVSQVTHLNGVSYDPTLMREHTPDSLPFGGGINPYSCYCLGDPINRSDPSGH-----1383  
 677 AANGWYLLNGVYKAYNPLMRPHSPFLSPFAEGGVNPPYCLGNPIALRDLPTGHDASGO 736  
 1384 -----LSWQ-----AWTGIGMGIA-----GLLLTATGMAIAAAGGI-----1416  
 737 TGLRLRPDEGALPMQOGGDMGWGVGIGVFTVLGVAATATLTATPVTGPTVLGI 796  
 1417 -----AAIASTSTTALAPG-----ALSVTSDITSIVSGALEDASPASSIILGWMS-----1462  
 797 SMTASAAAASVTSTGALIVGTALTAASTANTVAIVNN-----DQTAGEVGWGLGIAAV 851  
 1463 -----MGMGAAGLAESAIGKGTCLA-----THLGAFAEDENALLKSTSESSR 1505  
 852 PVGLVGFAGAGVVARAAVAAAANKAAVAAAGTIGVRSVSRIG-LAAAGARETTISSAASAR 909

RESULT 12  
 Q87VG6 PRELIMINARY; PRT; 1669 AA.  
 AC Q87VG6;  
 DT 01-JUN-2003 (TEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE YD repeat protein.  
 GN Pseudomonas syringae (pv. tomato).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID:323;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DC3000;  
 RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;  
 RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,  
 RA Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,  
 RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,  
 RA Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q.,  
 RA Khouri H.M., Pedorova N.B., Tran B., Russell D., Berry K.J.,  
 RA Utterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,  
 RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhouer S., Chatterjee A.K.,  
 RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Fang X.,  
 RA Bender C.L., White O., Fraser C.M., Collier A.;  
 RL "The complete genome sequence of the Arabidopsis and tomato pathogen  
 Pseudomonas syringae pv. tomato DC3000."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).  
 DR EMBL; AS016874; AA058398.1; -;  
 DR TIGR; PSPT04970; -;  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF05593; Rhs repeat; 5.  
 DR TIGRPFAM; TIGR01643; YD\_repeat\_2x; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 1669 AA; 184268 MW; BC668B6C1E3FB7EE CRC64;

Query Match 10.4%; Score 920.5; DB 2; Length 1669;  
 Best Local Similarity 25.9%; Pred. No. 8e-40;  
 Matches 440; Conservative 212; Mismatches 662; Indels 387; Gaps 84;

152 AFPLKVPKLLPAGHAIVDMFEATQPLRIYDDLDGHDIFLPLNLEYQGL---IKTI 208  
 72 AFQGLLPSEVTFPQGGTLSEWRNQGQFLKKI---SSDGNVI--LSSTYSPSDSTSTV 127  
 209 -LTLFPQGEKGYRTELRFNLRLQNLNHNFSGLNENPLTWSFGYPTGKNGILQWITSMT 267  
 128 EISVWPKTEAYTVRLSSYALTGLTRIEKNLKVSKLKYGYC---ADPTLDRVLNRIE 184  
 268 APGGLKETVYNNNGHHFPOSANLPVLPYVTLMKQVFGAPQAIQAEYSYTSNHYGG 327

185 EBDGSELVYTBREG--GMPPFYRQ--PPLPCVTLHSLPFGACQSNITDHYFYSGTNYLGF 241  
 328 GSGIWNKKLDNLYGLMTEYNTGSTRYKDKKEG-----DQIVIRBRTYNNYHL 378  
 242 SEPP--DAHQNELY-----YERLELRELVDEGVYQILRQNPDPHIVSTRHAFKNHL 292  
 379 LNSECKQNGYIQTETAYAIIGHNFDSPSQ--FQLPK--TKTETWRSANDSRYSEIT 434  
 293 QVRE-----DLQVQFAEKVISWBFANASPGKVFGLPTKITTDYTLDSHPNTERTTTV 346  
 435 ETTDESNGPLTKVTKDKTKKIISPSHWEYYPAG---EVD-----NC 476  
 347 QTLAVNNIGQLTKSI-----AVDGVTEWLYYPTDGGQGLDISLIAEKLKSLVTLTC 400  
 477 PPEYGFTRFVKKIIQTPYDSEFKDDP-----EKTIQRY-----SLIGS 516  
 401 PKVSEGY-----MPPVKVEYVHDPAKPSQIITAYAYARENPNQORSVLVPSVVV 452  
 517 QSHVTLKIEERHYSATQLLNSTLFOYNTDKSBLGRLLKQTECTKG-----ENGKTY--S 568  
 453 LTGVTL-----DITTMLPSLM-----EGRKNALIEQVRSIPDVATENTTAWKES 499  
 569 VVHKETTYTKODDTLOQSHSITTHDNFTIHR-----SQVRSRYTGRLFSDTDYK 616  
 500 VVQNSWLGWQNRNL--TTSMLYDDNPVSGRTVRAEAQKIISSRFSRLSGRLPSET-RD 556  
 617 DIVTQMSYDKLGRLLTR--TLNSGTPVANTL--TYDYEL--NNLQDDNRPFPVITTDVN 670  
 557 GLEPRYVHDSLGRVLRQERGTEAGAKADAVETTDYSITAEGLO-----VTVTEAE 608  
 671 GNQLRNEFDGAGRHSQCCLKOSD--GDGKPYTHIQYDEBQGRHHTSTYSYDITNGRQQT 729  
 609 -QVVRTLYDGLQRPVVAIKRTILPDSAPCVISRIEYDGPDATNQTLY--DYLPGGLRRTK 666  
 730 PDKVHLSMSKSYDNWQIANTHWSVG--VSEKITVDPITLTATKLOLSNNNVQTCKEVT 787  
 667 DARPEAAVDASKLAW--MADYTRFDAGLINEQVIGADSGAQLIRQLSGNLDKSHALLUE 725  
 788 TYPFS-----QQPIQITLDFEAGHLOSCHTLTRDGMVRVERKETDAIQCTIYQYDNNY 840  
 726 TLRPSTARDASDRIERT--FDERRLIKIRI-----SNTSEHC--IEYDELE 770  
 841 RVQITLDPGTTIVNRYKAPFSDTLITDIRVNGISLGQQTDFGLSRLTSQDQGRVWAT 900  
 771 RAVAIAPDGTGRTERKYHQLS--DYITQLNVGSTVLGTOKMTAAARQTTVGE---LTYE 824  
 901 YSAGNDQCPSTVITPDGQFIHQYQPELDDAVLOVASNEITQQ-----FSYNPVT- 950  
 825 FPGGS---ASTVVRPDKTLLESASVADGHTATLSINKKVHTQVVISQPNVLTVTDPVSV 881  
 951 -----GALLKVAEGQSITPIY--PSGRKKNENNDKMSYLMWTLRGLNGYDLDTCG 1003  
 882 PSAAEWSSLTSPQSLGITSITQTSFRSROAEMTRSLKG-----RLLTN--TAVDGR 932  
 1004 IQKISRDTGHRVTOIKDSISIKTTLNLYDDLNRHIGS---QVTDLATGMLTTTVPDGLN 1059  
 933 QMRVFRDYLDRVVRV---ILGELHYVLWSAFGEPLQRTVVNQASGERLDRVFTWDAFG 988  
 1060 REIGRKLCDSSGHTLDIOQSMKLTQOLANRIVKLVGLORTQOYSVDSNRNLNQYK--D 1117  
 989 QBIAREYTLNNKPLALLTNSYLANGQVSSKTLTREGVLQRTQRTGFSVADARDLNSVECTD 1048  
 1118 GAECPTDKYGHISVTFNQFTYDIYIGNITACHTTFADGTEDHATFKF---ANPTDPCOL-TE 1173  
 1049 VADWFQDQAGKSLKQSGYDEHLNLSSECSSTYADGSCIQYTYTDTVKNPTRLRSVKTE 1108  
 1174 VHH--THPMPDNIRLYDKAGRVINITDNHGNTEFTYDTLGLR---QNGQGSV---YGY 1226  
 1109 LRSGSQTSTQATLAYDANGN--QTTDESGR--LAVTPLGOLASVKDNGKLLTRYSY 1164  
 1227 DPLNRLVSKQTTDL--DCELYYRETMVNEV-----RNGEMIRLLRT-----GETII 1271  
 1165 DAFGLISQYIGATKHTCELLYDGTGQLTGEAWFDDANREFKRILFSEDMVQOCTIGETV- 1223



```
Db 286 EYEYASNYLVGASLGKAWNEDEDNINVMDDYTSSTEKLIVDNRE-----LVSISRI 341
QY 373 YNNYHLLTSECKQONGYIOTTETAYYAIIGHNFDSPQSQFOLPKTKTETWR-SADNSYRS 431
Db 342 YNSYLLISETTRQNSCEVIVETDYAKPGLSFDKQKQFOLPKKEKKTWRENSKNQCRS 401
QY 432 EITETTFDESIGNPLTKVVKDKTKQKIIISPTWHEYYPPAGEVDN-----CPPPEYGFTRFV 487
Db 402 EITTTTFDEGNLLTKIEPD-----GKTEYIYYDSKGETDKGIVLCPPEPENGFRFV 454
QY 488 KKIIOTPYDSEF-----KDDPEKFIQYRYSLSIGSQSHVTLKIEERHYSATQLLNSTLFOYN 543
Db 455 KTQIVTPADSEFYAPVQOTTYAYAQYPCIIAGSSLSYAVLQTOETLCSDDVL-----LLTIN 510
QY 544 TD 545
Db 511 TD 512

RESULT 15
ID O52883 PRELIMINARY; PRT; 526 AA.
AC O52883;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein orf 526.
GN Name=orf 526;
OS Cxiella burnetii.
OG Plasmid QPRS.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RA Lautenschlaeger S., Jaeger C., Willems H., Baljer G.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15898; CAA75844.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 526 AA; 60104 MW; 42D04FF8CC361CC CRC64;

Query Match 10.3%; Score 916.5; DB 2; Length 526;
Best Local Similarity 40.8%; Pred. No. 2.1e-40;
Matches 221; Conservative 73; Mismatches 203; Indels 45; Gaps 16;

QY 18 NEF-FTQANNFTSAVGGVDPRGLYNIOTILGHIVGNLGLPTLPLTLSSYSPANKTDI- 75
Db 2 NELPYTQATNFISAVQGVVDPRTGLFTVNNVLAELTGNLGNLGPDLFTLNYSPLSSTNIC 61
QY 76 GFGIGFNGLSVYDRKNSLLSLSTGENYKVIETDKTVKLQKKLDNLRPEKDKLKCNCYRI 135
Db 62 GFGIGCSVGISYDRKNKLLSSGERRYKIEDWNGVYVRQKINNFKFEK--IKNGYII 119
QY 136 IHKSGDIEVLGTGNNAFPLKVPKLLNPAGHAIYIDWNFEATQPLNRIYDDLDGHDIP 195
Db 120 KYKNGKTEYLYKYGNLNF---LPQKIFSTLGNWPLKLTWENRGQYVNLKKIEDAKD---V 172
QY 196 LLNLEYQGLIKLITLTPGQKEGYTELRFLNQLNSIHNFSIGNENPLTWSPGYTPICK 255
Db 173 LCKIDYQFSDWARITFWPGKTESYTFQLDFVNEYLYWVTNKTSTRE--LAWSFNYYDDVGA 230
QY 256 NGILGQWITSMTAPGLKETVNYNNQGHFPQSANLPVLPVYVTLMKQVPGAGOPAIQA 315
Db 231 GNFT---LTQVKSPTGLTETVNYQAGVM--RPDESCKPALPSVNYRQSPGMGQPDIVK 285
QY 316 EYSYTSHTNYVGGSN--GIWNKLDNLYG-LMTEYNYGSGTESRRYKDKEGHDOIVRIERT 372
Db 286 EYEYASNYLVGASLGKAWNEDEDNINVMDDYTSSTEKLIVDNRE-----LVSISRI 341
QY 373 YNNYHLLTSECKQONGYIOTTETAYYAIIGHNFDSPQSQFOLPKTKTETWR-SADNSYRS 431
Db 342 YNSYLLISETTRQNSCEVIVETDYAKPGLSFDKQKQFOLPKKEKKTWRENSKNQCRS 401
```

Search completed: February 9, 2005, 18:13:59  
Job time : 240 secs

```
QY 432 EITETTFDESIGNPLTKVVKDKTKQKIIISPTWHEYYPPAGEVDN-----CPPPEYGFTRFV 487
Db 402 EITTTTFDEGNLLTKIEPD-----GKTEYIYYDSKGETDKGIVLCPPEPENGFRFV 454
QY 488 KKIIOTPYDSEF-----KDDPEKFIQYRYSLSIGSQSHVTLKIEERHYSATQLLNSTLFOYN 543
Db 455 KTQIVTPADSEFYAPVQOTTYAYAQYPCIIAGSSLSYAVLQTOETLCSDDVL-----LLTIN 510
QY 544 TD 545
Db 511 TD 512
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2005, 17:59:43 ; Search time 58 Seconds  
(without alignments)  
2775.356 Million cell updates/sec

Title: US-09-889-874A-23  
Perfect score: 8879  
Sequence: 1 VIKFLKLFRLITMSDNEF.....PRKIILGRTEKTVKPTFRP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:\*  
1: Piri:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1214	13.7	709	2 S38241	hypothetical prote
2	912.5	10.3	528	2 S38242	hypothetical prote
3	432	4.9	2334	2 S32920	cell wall-associat
4	429.5	4.8	336	2 S38239	hypothetical prote
5	386	4.3	2167	2 AF1489	cell wall-associat
6	381	4.3	2183	2 T37218	hypothetical prote
7	365	4.1	1400	2 E90886	rhsE core protein
8	353.5	4.0	843	2 AB0539	Rhs-family protein
9	342	3.9	1397	2 C64805	rhsC protein precu
10	341.5	3.8	1404	2 E90658	rhsG core protein
11	341	3.8	1394	2 H91236	rhsH core protein
12	340.5	3.8	1411	2 E65145	rhsB protein precu
13	339	3.8	1426	2 H64780	rhsD protein precu
14	336.5	3.8	1404	2 E85509	hypothetical prote
15	331.5	3.7	1377	2 C63159	rhsA protein precu
16	331	3.7	1409	2 F91187	rhsA core protein
17	326	3.7	1399	2 A93720	rhsC core protein
18	322.5	3.6	1377	2 E86034	rhsA protein in rh
19	319	3.6	1397	2 A85570	rhsC protein in rh
20	316	3.6	1398	2 B85549	hypothetical prote
21	313	3.5	1398	2 H90698	rhsD core protein
22	297	3.3	1512	2 AH0439	probable membrane
23	280	3.2	1317	2 F83310	conserved hypotet
24	278	3.1	985	2 B86084	hypothetical prote
25	276.5	3.1	1438	2 AI0093	conserved hypotet
26	271.5	3.1	1354	2 AG0538	Rhs-family protein
27	266.5	3.0	2515	2 S47008	tenascin-like prot
28	248.5	2.8	1616	2 E90704	Rhs core protein w
29	245.5	2.8	1645	2 H85554	hypothetical prote

30 245 2.8 2825 2 T14271 Doc4 protein, stre  
31 230 2.6 794 2 T36972 probable membrane  
32 230 2.6 849 2 T46253 hypothetical prote  
33 220.5 2.5 1185 2 A42404 collagen adhesin -  
34 219.5 2.5 1962 2 A32634 lactocepin (EC 3.4  
35 218.5 2.5 2406 2 A54148 odz protein - frui  
36 217 2.4 356 2 T37136 hypothetical prote  
37 216.5 2.4 2894 2 C64474 hypothetical prote  
38 215.5 2.4 2044 2 AB1180 probable peptidogl  
39 213 2.4 1487 2 AG2560 hypothetical prote  
40 211 2.4 3283 2 AC1018 large repetitive p  
41 208 2.3 2703 2 H81193 hemagglutinin/hemo  
42 207 2.3 2893 2 A64556 toxin-like outer m  
43 206 2.3 1902 2 S06997 lactocepin (EC 3.4  
44 203.5 2.3 1366 2 S57664 IGA-specific metal  
45 202.5 2.3 1349 2 A11476 cell surface prote

RESULT 1  
S38241  
hypothetical protein - Coxiella burnetii  
C:Species: Coxiella burnetii  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
C:Accession: S38241  
R:Thiele, D.; Willems, H.; Haas, M.; Krauss, H.  
submitted to the EMBL Data Library, October 1993  
A:Reference number: S38215  
A:Accession: S38241  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-709 <THI>  
A:Cross-references: UNIPROT:Q45948; EMBL:X75356; NID:G407370; PIDN:CAAS3129.1; PID:G4073

Query Match 13.7%; Score 1214; DB 2; Length 709;  
Best Local Similarity 40.8%; Pred. No. 5.6e-59;  
Matches 287; Conservative 101; Mismatches 280; Indels 36; Gaps 16;

QY 554 KQTECTKGNGKTVSVVHKFTYTKQDDTLQOSHSTTTHDNFTIHRSVQRVSRVYGRFLFSDT 613  
Db 3 RRAEVLTSKREKKYQNQTTFATISSQAEHLLOKIDFTGGDGKITSIRQSRVSGHLLSST 62  
QY 614 DTQDITVQMSYDKLGRLLTRTLN-SGTPYANTLTVDYELNNLQDDNRPFPFVITTTDVNGN 672  
Db 63 DELGNVTQVEYDELGRLLTQTVNASSTYASTRYSYSLTDARGKVTAIKTTVTDPKGN 122  
QY 673 QLRNEFDGAGRHVSQCLKSDG-----DGKFYTHIQYDEQGRHHTSTYSYDLNGRQQ 727  
Db 123 QLRTYDGLGRNLKQERLDKAAVSQKTGTWTYTHQQYDGLGRESKTTQDVL-----R 177  
QY 728 TDPDKVH-----LSMSK--SYDNWQIANTHSYGVSEKITVDPTLTATKQLOSNNVQ 781  
Db 178 LDSEVGHAGSVLSTSKVMHYDSWGNHLTVFSDGQERSVYDPIFRRAT--LQPSGSGOK 235  
QY 782 TGKEVTTTPSQOPIQITLDFEAGHLQSCHTLTRGDWRVRKETDAIGQCTTYQYDNYNR 841  
Db 236 LGQQLTEYNLAGLPKVTQYDSQGTQSAHVEYDGLQLRKETDELQITLYEYDHFGR 295  
QY 842 VIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFFDGLSKRLTSQDQGRVWAYTY 901  
Db 296 VTQTTLPENTIIQKSYAPHSTASLTIGSVNNFSGNQTFFDSLERLTETTSGRTSAFSY 355  
QY 902 SAGNDQCSTVITPQGFHYQYQPELDDAVLOVASNITQOFSYNPVTGALLKAV-REG 960  
Db 356 ENAS-SVPAVAVTAPTGETVSVEYLKELGNVAKKISAPEILOQYDWDALTGAMTSATQAG 414  
QY 961 QSLTPIYVYGRKME-NIND---MKMSYLTWLTGLENGVYTDLTGTTQIKSRDTHGRVT 1016  
Db 415 MIRQMTYPSGLLKKNETSNPDGAOKSFATYYSLAGAPQSYYTDFVGTQRYDYDEHGRRI 474  
QY 1017 QIKDSSIKTTLNYYDNLNRHIGSQVTDLATGHHMLTTTVFDFGLNLRIGRKLCDSSGHTLDI 1076

Db 475 GIEDNDIKVSLDYDAFRGRTKQQAIDDKTGAVLSTLTLYDDLNREIKREISASGQSVLVI 534  
Qy 1077 QQSWLKTOQLANRIYKLVGLVLTQTEQSYSDNSNRLNOYKCDGAECPYTKYGHISVITQFT 1136  
Db 535 EOTYQRNHLKERTIYQRTGRTLLKRMFAYDSNRLIEYTCNGEARPDQPKYKAIHQRTFS 594  
Qy 1137 YDIYGNITACHTTFADGTEDHATFFANPTDPCQLTEVHTHPDMPDNIRLKYDKAGRVI 1196  
Db 595 YDALGNWTKQTFDFSGG-RNTATYIY-SAIDPTQLKVNNDHSDYPKETILEYDKAGRWI 652  
Qy 1197 NITDNGHNTFNTYDTLGRLO-NG---QGSYGYVDPLNRLVSQ 1235  
Db 653 R-DEAGRT--LRYDALGRLOQVNGAGAKGGQYAYDVNLTVLSQ 692

RESULT 2  
S38242  
hypotheical protein - Cxiella burnetii  
C:Species: Cxiella burnetii  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
C:Accession: S38242  
R:Thiele, D.; Willems, H.; Haas, M.; Krauss, H.  
submitted to the EMBL Data Library, October 1993  
A:Reference number: S38215  
A:Accession: S38242  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-528 <THI>  
A:Cross-references: UNIPROT:Q45949; EMBL:X75356; NID:g407370; PIDN:CAA53130.1; PID:g40737  
C:Superfamily: Cxiella burnetii hypotheical protein

Query Match 10.3%; Score 912.5; DB 2; Length 528;  
Best Local Similarity 40.8%; Pred. No. 1.4e-42;  
Matches 222; Conservative 72; Mismatches 203; Indels 47; Gaps 17;

Qy 18 NEF-FTOANNFTSAVSGVDPRTGLYNIQITIGHIVGNGLGPTLPLTLTSLSPLNKTDI- 75  
Db 2 NELPYQTATNFTISAQGVGVDPRGTGLFTVNNVLAELTGNDNLGDPDLFTLNVSHLSTNIC 61  
Qy 76 GEGIGFNGLSVYDRNSLLSSTGENTKYIETDKTVKLOQKLDNLNLFKDLKENCYRI 135  
Db 62 GFGICGSVGISYDRNKKLLLSGGERYKTEDWDSVYRQKINNFKFEK--IKNGYII 119  
Qy 136 LHKSGDIEVLTG--FNNNAFDLKVPKLLNPAGHAIYIDWNFEATQPLNRIYDLDLGDH 193  
Db 120 KYKNGTEVLNRYKYGDNLF---LPQKIFSLPGWPLKUTWENRGQYVNLTKIEBAXO--- 173  
Qy 194 IPLLNLLEYQGLIKTILTLFPQKQEGYRTLRPLNRLQNSIHNFSIGNENPLTWFGYTPPI 253  
Db 174 -VLCKIDYQFSDWARITFWPGKTESYTFQLDFVNEVLYVWTKTSRE--LVWSENYDDV 230  
Qy 254 GRNGILGQWITSMTAPGLKETVYNNNNQGHPPQSANLPLVPLVYTLMKQVPGAGQPAI 313  
Db 231 GAGNFT---LTQVKSPTGLTETVNYQAGVM--RFPDESQKALPSVYNYRQSPGQGPDI 285  
Qy 314 QAEYSYTHNVYGGSN--GIWNKKLDNLG-LMTENYNGSTESRYKDKEGHQIIVRIE 370  
Db 286 VKBEYTVSNILGYGASLKAWNEDNINVMDDYTSSTKELIVDNR-----LVVIS 341  
Qy 371 RTYNHYHLLTSECKQONGYIQTETAYYAIIGHNFDSPQFQPLPKTKTETWR--SADNSY 429  
Db 342 RIYNSYLLISSETTRQNSCEVIVETDYAKPGLSPDKQPKQFQPLPKBEKKTWRENSKQC 401  
Qy 430 RSEITETFDSEGNPLTKVIKDKTKQKIIISPTHWYPPAGEVDN-----CPPBPYGBTR 485  
Db 402 RSEITTTTFDEGNLLTKIEPD-----GKTEVIYVYDSKGETDKGIVLCPPEPNGFVR 454  
Qy 486 FVKKTIQTPYDSEF---KDDPEKFIQYRYSIGSQSHVTLKIBERHYSATQLLNSTLFPQ 541  
Db 455 FVKTIQVTPADSEFYAPVQQTYYAYAPCPICAGSSLSYAVLQETCLSDVDL-----LLT 510  
Qy 542 YNTD 545

Db 511 INTD 514

RESULT 3  
S32920  
cell wall-associated protein precursor wapa [similarity] - Bacillus subtilis  
C:Species: Bacillus subtilis  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C:Accession: S32920; E69730; T47101  
R:Forster, S.J.  
Mol. Microbiol. 8, 299-310, 1993  
A:Title: Molecular analysis of three major wall-associated proteins of Bacillus subtilis  
protein.  
A:Reference number: S32919; MUID:93302506; PMID:8316082  
A:Accession: S32920  
A:Molecule type: DNA  
A:Residues: 1-2334 <FOS>  
A:Cross-references: UNIPROT:Q07833; GB:L05634; NID:g304177; PIDN:AAA22883.1; PID:g304179  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emmerson, P.F.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
A:Authors: Schlicht, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowaka, A.; Seror  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.P.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: E69730  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2334 <KUN>  
A:Cross-references: GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CAB15959.1; PID:g2636469  
R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.  
Microbiology 141, 337-343, 1995  
A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contai  
A:Reference number: 224350; MUID:95219088; PMID:7704263  
A:Accession: T47101  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2334 <YOS>  
A:Cross-references: EMBL:D31856; NID:g603765; PIDN:BAA06656.1; PID:g603782  
A:Experimental source: strain BGSC1A  
C:Genetics:  
A:Gene: wapa; N17G  
C:Superfamily: cell wall-associated protein wapa

Query Match 4.9%; Score 432; DB 2; Length 2334;  
Best Local Similarity 20.4%; Pred. No. 5.6e-15;  
Matches 358; Conservative 226; Mismatches 613; Indels 560; Gaps 80;

Qy 44 IQITLGHIVG-NGNL-----GPTLPLTSLSPLNKTDIGFIGNFGLSVYDR 90  
Db 896 IDIPSGQLNGATGNVIVNEEDLSIDGRGPGGLGSLRTYNSLDSSDHLFGQGW-----YADA 950  
Qy 91 KNSLLSLSTGENYKVIETDKTVKLOQKLDN-----LRFKDLKENCYRIHKSGDIEVL 145  
Db 951 ETSVISTDGGAMY--IDEDATTHRTKKADGTQPTGTVVLELTETADQDFILTKD---- 1004  
Qy 146 TGFNNAFDLKVPKLL-----NPAGHAIYIDWN-----FEATQPLNRIYDLDLGDH-- 192  
Db 1005 ---OTNAYFKKGGKLOKQVGDGNNAVYVNDKQQLTATDASGRKLTFTYDSE-NGHVT 1060  
Qy 193 -----DIPLLNLEY--OGLI-----KTLTLFPG 214



Db 1061 SITGPKKKVITYSYENDLLKKVYTDGTVTSYDYDSEGLVKQYSANSTEAKPVTEY-- 1118  
QY 215 QKEGRTRELFNRQLNSIHNF-----LGNENPLTWSFGYTPIGKNGILQOMTSM 267  
Db 1119 QYSHRLEKAIKAKKETYVYSYDAADKTLMTQPNRKYQYGYNEAGNP-----IQVID 1172  
QY 268 APGGLKETVN--YSNN-----QGHFFPQSANLPVLPVYTLMKQVPGAG 309  
Db 1173 DAEGUKITNTKYEGNNVVEDVDNDVGTGKATESYQYDKGN-----VTSVKDAYGT- 1225  
QY 310 QPAIOAEYSYTHNYVGGSGNINWKNLNDLGLMTEVNGYSGESRRYKDKEGHQDIQVRI 369  
Db 1226 -----ETYEYNNKNDV-----TKMDTSGNVTDIAYDGLDAVSETDQSGKSSRAV 1271  
QY 370 ERTYNNYHLLTSECKQONGYIOTTETAYYAIIGHNFDSPQSFQPLPKTKTETWRSADNSY 429  
Db 1272 YDKYGNQIQSSKDLGASTNLK-----DGSFEAQKSGWNLTASKD----- 1311  
QY 430 RSEITETTFDESGNPLTKVKKTKQKIIISPS-----THWEYYPAGEVDNCPPEPYGTR 485  
Db 1312 RKKIS-VIADKSG-----VLGSKALEVLQSSTAGTDHGYSSATQTVELEPNITTYTLG 1365  
QY 486 FVKKII---OTPYDSEFKDDPKFIQ---RYSLIGSQSHVTLKTEERHYSATQLNS-- 537  
Db 1366 KIKTDLAKSRAVFNIDLRDKQRRIQWIHNEYSALAGKNDWT-----KQITFTTANAK 1421  
QY 538 ---TLFQVNTDKSELGRL-LKQTECTKGENGKTSYVWHKFTYTKQDDTLQOS-HSITTHD 592  
Db 1422 AVVMEVDHKDKGKAWFDEVQLEKEGVSSYPVQNSFTSATENWNVSGASVDSEE 1481  
QY 593 NPTIHRQVRSRYTGRFLPSDPTKDIIVTQMSYDKLGRLLTRT-----L 635  
Db 1482 GFNDVSLKAARTSASQAGSVTKQTVVLQGSANDKPVVLTGLTMSKASSVKFTDEKDYSL 1541  
QY 636 NSGTPYANTLTVDYEL---NNLQDNRPPFVI----- 664  
Db 1542 QANVTYADGSTGIYNAKPSGTQEWNRAAVVPKTKPINKVDISILFQKSATGTWVFDDI 1601  
QY 665 -----TTTVDNGNOLRNEFCAGRHVSQCLKSDGDKGFYTIHTHQYDEQGRHHT 714  
Db 1602 RLIEGSLTKSYDSNGNVVTKDEELGYATS---TDYDETK---KTSETDAKEXTT 1654  
QY 715 STY---SDYLTNGRQOTDPDKVHLSMSKSYDNWG-QIANT-----HWSYGVSEKI-- 760  
Db 1655 YTYDAQDLTNMTLSNGTSILH-----SYDKEGNEVSKTIRAGADQTVKFEYDVMGKLVK 1709  
QY 761 TVDPITLTATKQLQNSNNVQT-----GKEVTTYPSQQPIQITLFDKAGHLOSCHTLTRD 816  
Db 1710 TTDPILGNVLASEYDANSNLTKTISPNGNEV-----SLSYD 1744  
QY 817 GWRVVRKETDAIGQCTIYOYDNYNRVQITLPGDTIVNRKYAPFSTDTLITDIRVNGISL 876  
Db 1745 GTRVRSKSYNGTEKIFYTDKNGN-----ETSVNKEQN-----TT 1781  
QY 877 GQCTFDGLSRLTQSDGGRVWATYTSAGNDQCFSTVITPDGQFIHYQYQPELDDAVLQA 936  
Db 1782 KKRTFDKNRLFELTDRCGSSQWTYPDSQDLKTP-----SWIH-----G 1821  
QY 937 SNEITQCFSPNVPVGTALLKAVAGOSLPTIYYPGRLKMENTNDMKMSYMLTLGLENG 996  
Db 1822 DQKGINQFTYN-----KLDQMIEMKDSSTSSYSFYDEN- 1854  
QY 997 YTDLTCTIQKISRDTGRTVQIKDSSIKTTLNLYDDLNHRHIGSQVTDLATGHMLTTTVEFD 1056  
Db 1855 -----GNVQ-----TFITNGGGTSFSYDERNLVSSLIHGDKNXGDLILTESYEX- 1898  
QY 1057 GLNRBIGRKLCDSSGHTLIDIOOSWLKTQOLANRIVKLGNVLQRTQYSYDSNRNLNOYKC 1116  
Db 1899 -----DANGNRTTINS-----ASGV-----QYEGKLNQLVK--- 1927  
QY 1117 DGAECPTDKYHSIVTQNFYDIYGNITACHTTFADGTED--HATFKFANPTDPCQLTEV 1174  
Db 1928 -----ETHEDGTVI---EYTDGFGNRKTV--TTIKDGSKTVNASFNIMN-----QLTKV 1973

QY 1175 H-----HTHP-DMPDNI-----RLKYDKAGRVINITDHHG 1203  
Db 1974 NDESISYDKNGNRTSDGKFTYTWDAEDNLTAVTKKGEDKPFATYKYDEKGNRIQKTVN-G 2032  
QY 1204 NTEFTYDTLGRQLQNGQSVGYDPLNRLVSQKTDLDCELY-YRETMLVNEVRGEMIR 1262  
Db 2033 KVTNYFYDG-----DSLNVLYETDADNNVTKSYTYGD-----SQLLUS 2070  
QY 1263 LLRTGETTIAQORASKVLLTGTDSQQSVILTSDKONLSQEAYSAYGK-HKSTANDA---S 1318  
Db 2071 YTEGKKFYHYNAHGDIIAISDSGKTV-----AKYQYDANGNPTKTEASDEVKDN 2122  
QY 1319 ILVNGERADPVSGVTHLNGYRSYDPTLMRPHT--PDSLSPFGAGGINPYSYCLGDPIN 1376  
Db 2123 RYRYAGYQYDEBTGLYLMARY--YEPRNGVFLSLDPPDPSGDSGLDQNGYAYGNPNPVM 2180  
QY 1377 RSDPSHLSSQAWTIGMGIAGLLLTIATGMAIAAAG---GIAAAASTSTTALAPGA 1432  
Db 2181 NVDPDGH--W-VMLVNVNAGFA-----AYDGKYAYKSGKMGKAAWAAASNFGPKIPKG 2231  
QY 1433 LSVTSDITSIVSGALEDASPKASSILGWVSMGM-----GAAGLAESAIKGGTKLATHLAF 1488  
Db 2232 ASRAYKFTK-----KAVKITGHTHGLNQSIGRNG-----GRGVNLRAKLNA- 2273  
QY 1489 AEDGENALLKSTSESSRIKWGVTRSLDRE--IVRNEEQGVINKHRSRGYTDNPMGKGEQAI 1546  
Db 2274 -----VRSPKVKIQPNGATKYVGKATVVLNKRKGVITAYG-----SS 2312  
QY 1547 LVHGDQKGFVHTGK 1563  
Db 2313 RAKSGKHVFHTHGK 2329

## RESULT 4

S38239

hypothetical protein - Coxiella burnetii

C:Species: Coxiella burnetii

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004

C:Accession: S38239

R.Thiele, D.; Willems, H.; Haas, M.; Krauss, H.

submitted to the EMBL Data Library, October 1993

A:Reference number: S38215

A:Accession: S38239

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-336 &lt;THI&gt;

A:Cross-references: UNIPROT:Q45946; EMBL:X75356; NID:G407370; PIDN:CAA53127.1; PID:G4073

## Query Match

Best Local Similarity 4.8%; Score 429.5; DB 2; Length 336;

Matches 121; Conservative 48; Mismatches 91; Indels 93; Gaps 13;

QY 1348 MRFTPTDLSLSPFGAGGINPYCYLGDPINRSDPSGHLSSQAWTGMGIAGLLLTATGG 1407

Db 1 MRINCPDSSWSPFGAGGINPAYCDDPINRVDPNGLHSQWAEIDIGLVGLVLAFTAG 60

QY 1408 MAIAAAGIAAIAASTTALAFNALSVTSDITSIVSGALEDASPKASSILGWVSMGMA 1467

Db 61 TSAIAAGIAAIAESASAIISLVGTLGVAADVASIASGALEDANPOASATLGLWSLGLGG 120

QY 1468 AGLAE--SAIKGGTKLATHLGAFAEDG-----ENALLKSTSESSRIKWGV--RS 1513

Db 121 PGAVSGLATARAAGKL---ISGLAKGGKTRSPVQGISYRSLSRGDPLRGPPHQS 177

QY 1514 LDREIVRNEE-----GOVIKDHSGYTDNF-----MGK 1541

Db 178 LSRVTVAPESMRPAGLNWHKVSQKSLGYQHVFCADREIFGYETREIPEFRRRPSITK 237

QY 1542 GEQAIL-----VHGDKDGFVHTG-----NKHNGKGPYTRHTPQLVDY 1581

Db 238 RDIVILSGTHGRVHGDN---WTSQGLRRPDLERAFYIEDVQNYKG-----QLNGR 285





Db 2108 TYSGRARISS---INKEGVSITLKFSAWNGSD---WRSATHVVPWSNP 2150

RESULT 7

E90886

RhsE core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004

C;Accession: E90886

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A99629; PMID:21156231; PMID:11258796

A;Accession: E90886

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1400 <HAY>

A;Cross-references: UNIPROT:O8X2F8; GB:BA000007; PIDN:BA035484.1; PID:gl3361527; GSPDB:C

A;Experimental source: strain O157:H7, substrain RMD 0509952

C;Genetics:

A;Gene: ECS2061

C;Superfamily: rhesF protein

Query Match 4.1%; Score 365; DB 2; Length 1400;

Best Local Similarity 20.9%; Pred. No. 1.1e-11;

Matches 284; Conservative 154; Mismatches 467; Indels 456; Gaps 65;

QY 418 KTEW-READNSYRSEITTTTDFESGNPLTKVI-----KDKTKIISPTSHWEY 466

Db 218 RLTVRREAAGDLAGEITGVT-DGAGREFRLVLTQAQRAEARKQHTASLSSPDT--- 272

QY 467 YPAGEVNCPEPGFTRFKVKKIITPYDSEPKDPEKFTQYRYSLLGSQSHVTLKIEE 526

Db 273 -----PRPLSDSAFPDPLPGTEY-----GPDGRGIR----- 297

QY 527 RHYSATOLLNSTLFQYNTDKSELGRLLKQTECKENGKTVSV-----VHKFTYTKQD 579

Db 298 ---LSAVLTHDPAYPESILPGAPLARY-----TYTEAGELLAVDRSNTQVRAFTYDAQH 349

QY 580 DTLOQSHSITTHDNFTIHRSQVRSYRTGLFSDDTKDIVTQMSYDKLGRLLTRFLNSGT 639

Db 350 PGRMVAH-----RYAGK-----PEMYRYDDTGRVVEQLNPAGL 383

QY 640 PYANTLYDYELNNLODNRPFVITTD-VNGQLRNEFDGAGRHVSQCLKSDGCKGF 698

Db 384 SY-----RYQYE-----ITVTDLSNRREVLTGEGAG--LKRVRVKELADG-- 425

QY 699 YTIHTQQYDEQGRHHTSYSDYLTNGR--QQTDPDKVHLSMSKSYDNNWQIANTHWSYGV 756

Db 426 -----SVTHSGYDAAGRLTAQTD-----AAGRTEYGL 453

QY 757 SEKITVDPITLPAKLOSNMNVQTEKVTYTPSQOPIQITLDFEAGHLQS-----CHT 812

Db 454 N--VVSQDITDITPDGRETREKYNDGNQLTAVVSPDGLSRAYDEPGLVSETSRGCD 511

QY 813 LTRDGDWRVKE-----TDAIGQCTIYQYDYNRVIIQITPDGTIVNRKYAPFSTDLIT 867

Db 512 VIRAYDNPHELPAITTDATGSTRQMTWSRYGQLLAFTDCSGYQTRYEYDFRFGQMTAVH 571

QY 868 DIRVNGISLGOQTFDGLSRLTSQDG-GRVMAITYSAGNDQCPSTVITPDGQFIHYQOP 926

Db 572 ---REGIS-RYRYRNNRGLTSVKDAQGHEFYENNAAGDL--TAVITPDGNSRTQY-- 624

QY 927 ELDDA---VLQVASNEITQQFSYNPVTGALLKAVAGOSLPIFYYPSPGLKMNENDMKK 983

Db 625 ---DAWKAVSITQGLTRSMYED-LAGRITTLTNWNGSRSEFTYDA----- 667

QY 984 MSYLWTLRGLNGYTDLTGCTIQKISRDTHGRVTOIKDSIKITLNYDDLNRHIG----- 1037

Db 668 LDRLVQQRGFD-----GRTQRYHYDLTGKLTQSEDEGLVTLWHYDSDRLTHRTVNGE 720

QY 1038 -----SQVTDLATGMLTFTVTFDGLNRIRGRKLCDSGHTLDI-----QQS 1079

Db 721 PAEQWQYDEHGWLTEISHLSEGHQVAVHYGYDDKGRLAGERQTVNPNPETGELLWOHETE 780

QY 1080 WLKTOQLANRIVKLVGLQTEQSYDS-----RRL-----N 1112

Db 781 AYNSQGLANRVP--DSLPRVEMLTYSGLAGMKLGCTPLVETFRDLRHRETIVRSFGNN 838

QY 1113 QYKCDGAECPDKYKH-----SIV-----TQNFYDIYGNIT 1144

Db 839 AYELTSYTFPA---CHLQSQRLNSQVDRDYDNDNGDLVRISGPRQTEWYGYGATGELE 895

QY 1145 ACHTTFADGTEDHATKFAFNPDPF--QLTEVHHTHPD-----MPDNRL--KY 1189

Db 896 SVRTLASD---LDIRIPATDPAGNRLPD-PELHPDSTLTAMPDN-RIAEADAHYVVRH 948

QY 1190 DKAGRVINITD-----NHGNTENFTYDTLGR-----QNGQSV---YGYDEL-- 1229

Db 949 DEYGLRTEKTDRI PAGVIRTDDETHHYDQSRQLVFFYTRIQGEPLVESRYLDPLGR 1008

QY 1230 -----NRLVSQKTDTLDCELYR----- 1247

Db 1009 RMAKRVRRERDLTGWMSLSRKPEVTWYGVGDRLTIVQDITRIQTVYEPGSETPLRV 1068

QY 1248 -----ETMLVNEVRNG-----EMIRLLTGTETIIAQORAS----- 1277

Db 1069 ETENGEREKAQRRLAETLQOEGSENGHVFPAPAEVLRLDLRLLEEIRADRVSSESAWL 1128

QY 1278 -----KVLITGTSQ--QSVILTSQKQLSQEAYSACKHKST 1313

Db 1129 AQCLTVEQLARQVEPYTPARKVHYCHDRHGLPLALISEDGNATWARGEDYDEGNQUNE 1188

QY 1314 ANDASI---LGYNGERADPVSVTHLNGYRSYDPTLMRPHTPDSLSGPFQ-AGGINPYSY 1369

Db 1189 ENPYLHQPLRQCOHDESGLYNRY--YDPLQGRYITQD---PIGLAGGNWLYNY 1243

QY 1370 CLGDPINRPSFG-----HLSWQAWTIGMGIAGLLT-----IATG---GMAIAAAGGI 1416

Db 1244 PL-NPIIRMDPLGLYNYQLLYDVWHDDSVYGTSDIDTSGDGLISLGHAGLGVAFAKKK 1302

QY 1417 AAATASTTALARGALSIVTSDITSIVSGALEDASPKA--SSILGWYSMGACAGLAESA 1474

Db 1303 GEMLSDICIVATACCHAGIGGINAAITYSYKSLPTSGVSNVSG-VTVGGVGG----- 1356

QY 1475 IKGQTKLATHLG-AFAEDGENALKSTSESRIRKGVTRSL 1514

Db 1357 -----HPAYTVVDVNDP--ESSTESVGICAGVDASV 1386

RESULT 8

AB0539

Rhs-family protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain A;Species: Salmonella enterica subsp. enterica serovar Typhi

A;Note: this species has also been called Salmonella typhi

C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C;Accession: AB0539

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar AB0502; MUID:21534947; PMID:11677608

A;Reference number: AB0502; MUID:21534947; PMID:11677608

A;Accession: AB0539

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-843 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD08754.1; PID:gl6501575; GSPDB:GN00176

C;Genetics:

A;Gene: STY0324

Query Match 4.0%; Score 353.5; DB 2; Length 843;

Best Local Similarity 22.8%; Pred. No. 2.1e-11;







QY 1230 -----NRLVSQKDTLDCELYR-----ETMLVNEVRNGEMI 1261  
DB 1019 RERDLTGWMSLSRKRPETWYGDGRLTTVQQTTRIQTIVYQPGSTPLRLRIETENGEOA 1078  
QY 1262 RLLR-----TGETIIAQ-----QRASKVLLTGTDSQOS-----VILTSKQNLIS 1300  
DB 1079 KARHRSIAELVQEDTGTPLPAELAVMLGRLELRGSGVSSESQWLQAQGLTASQMGQA 1138  
QY 1301 QEA-----YSAYGKHKSSTANDASI-----LGYN 1323  
DB 1139 LEAGVIPERKLHLYHCDQRLGLPLGLISPGRETALTAEYDEWGNLLSETSAQPLQOSLRPP 1198  
QY 1324 GERADPVSGVTHLNGYRSDYDPLMFHTPDSLSPEG-AGGINPYSCYCLGDPINSDPSG 1382  
DB 1199 GQYDEESGLYNNRY--YDPLQGRYITOD---PIGLEGGWNLVQYPL-NPIEHIDPLG 1252

RESULT 11  
H91236  
RhaH core protein with extension [imported] - Escherichia coli (strain O157:H7, substra  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C:Accession: H91236  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Iehii, K.; Yokoyama, K.; Han, C.G.  
gaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: H91236  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1394 <HAY>  
A:CROSS-references: UNIPROT:Q8X385; GB:BA000007; PIDN:BA838287.1; PID:gl3364340; GSPDB:Q  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Gene: ECs4864  
C:Superfamily: rhes protein

Query Match 3.8%; Score 341; DB 2; Length 1394;  
Best Local Similarity 21.5%; Pred. No. 2.4e-10;  
Matches 260; Conservative 156; Mismatches 404; Indels 392; Gaps 63;

QY 418 KTEW-RSADNYSREIHTTTFDESGNPLTKVI-----KDKTKIIS-----PS----- 461  
DB 218 RTQTFHREAAGFSGEITGT-DGAGRHLVLTQAQRAEARQAISGGTEPSAFPD 276  
QY 462 -----THWEYYPAGEVNDCCPEP---YGET-----RFFK 488  
DB 277 LPGYTEYGRDNGIRLSAVWLTHDPEY-----ENLPAFLVRYGWTTPRGELAVVDRSGK 331  
QY 489 KIIQTPYDSEPK-----DDPEKFIQYRSLIGSQSHVTLKIEERHYSATQLLNSTL 539  
DB 332 QVRSFTYDDKYGRVVAHRTGRPE--IRYRD---SDGRVTEQLNPAGLSYT----- 379  
QY 540 FOYNTDKSELGLRLKQTEC--TKGNGKTYSVVHKFTYTKQDDTLQQSHSIITHNFTIH 597  
DB 380 YQYKDRITITDLSNRREVLHTQSGG-LKRVVK-----EHADGSVT 421  
QY 598 RSQVRSRYTGLFSDTDKDIIVTQMSYKLCGLLFTLNSGTPYANTLYDYELANLQDD 657  
DB 422 QSQFDA--VGRURAQTDAAGRTEYSPDVVTGLITRIT---TPDGRASAFYF----- 468  
QY 658 NRPPFVITITVDVNGNQLRNEFDGAGRHSVQCLKSD-----GDGKFYTHITQYVQOGRH 712  
DB 469 NHHSQLTSATGPDGLIEIRYDEWGLIETAPDGDITRYRDNPHSDLPACATEATGSR 528  
QY 713 HTSTYSYLTNCRQQTDPDKVHLSMSKSYDNNWQIANTHWSYGVSEKIVDPPI-TLTATK 771  
DB 529 KTWMTWSRY---QQLSFTDCSGVYTRYDHRFQMTAVHREGLSQRAYDSRGQLIAVK 585  
QY 772 QLOSNNSNVQTKVETTPSQOPIQITLFDAGHLQSCHTLTRGWDVRKXETAIGA- 830  
DB 586 DTQGHETRYE-----YNAAGDLTT--VIAPDG-SRNGTQYDAWGKA 623

QY 831 CTI-----YOYNNYNRVIOITLPDGTIVNRKXVAPPSTDTLITDIRVNGISLGQOTF- 881  
DB 624 ICTTQGLTRSMYDAAGRVIRLTSENGS-----HTTFRYDLRLIOETGDFGRQRYH 678  
QY 882 -DGLSRLTOSQDGRVWYATYSAGNDQCPSTVITPDGQFIHYOQPELDAVLQVASNEI 940  
DB 679 HDLTGKLIRSEDEGLV-----THHYD-EADRLTHRTVKGET 714  
QY 941 TQOFSYNPVTGAL--LKAVAEGQSITPIYV---PSGRL--KMNINDMKMSYLM----- 988  
DB 715 AERQWYDE-RGWLTIDISHISEGRVT-VHYGYDEKRGRLTGERQTVHHHPQTEALLWQHEIR 772  
QY 989 ---TJRLGNGVYDTLGTIQIKISRDTHGRVTOIKDSSIKTTLNYDDLNHRHISQVTDLAT 1045  
DB 773 HAYNAQGLAN--RCIPDLSLPAVEWLTYG-----SWLAGMKL 807  
QY 1046 GHMLTITTFEF--DGLNREIGRKLDCSSGHTLIDIQOSWLKTOQLANRIVKLVGLQRTQY 1103  
DB 808 GD--TPLVDFTDRDLRHLRKTLLRFGRYELTTATPAGQOSQHLS-----LQYDRDY 857  
QY 1104 SYDSERNLNQYKCDGAECPDKYGHISIVTQNFYDIYGNITACHTTTFADGTBDHATFKFA 1163  
DB 858 TWNDGELIR-----ISSPRQ-----TRSYSDSGELTGVHTTAA-----NLDIRIP 900  
QY 1164 NPTDFC--OLTEVHHTHPD-----MPDN-----IRLKYDKAGRVINITD----- 1200  
DB 901 VATDPAGNRLPD-PELHPDSTLSMMPDNRIARDAHLYRYDRHGRLLTEKTDLIPEGVIRT 959  
QY 1201 NHGNTENTYDTLGLRLQNGQ-----GSYGVYDPLNRLYSOKTDTLDCEL----- 1244  
DB 960 DDERTHRYHYSQHRLVHTYRTQYBEPLVESRYLYDPLGRRVAKVRWRERDLTGWMSLS 1019  
QY 1245 -----Y-----RETMLVNE-----VRNGEMIRLLR-----TGETIIAQORA----- 1276  
DB 1020 RKPQVTWGDGRLTITQNDTRIQTIVYQPGSTPLIRIVETATGSLAKTQRSLADAQ 1079  
QY 1277 -----SKVLTGTSQQSVILTS----- 1294  
DB 1080 QSGEDGGVFPVPLVQMLDRLESEILLADRVSERRWLASCGLTVAQMOSQMDPVYTP 1139  
QY 1295 -----DKQNL-----SQEA-----YSAYGKHKSSTAND-----ASILYNGERADPV 1330  
DB 1140 ARKIHLYCHDRGLPLALISKEGATEWCAEYDEWGNLLNEENPHQLQLIRLPGQYDSE 1199  
QY 1331 SGVTHLNGYRSYDPTLMRFHTPDSLSPEG-AGGINPYSCYCLGDPINRSDPSGHLSWQAW 1389  
DB 1200 SGLYY--NRHRYDPLQGRYITOD---PIGLKGGWNLVITYPL-SPVNGMDPLGLYEFKSK 1253  
QY 1390 TGIGMGIAGLLL 1401  
DB 1254 NIDDIGIFALAM 1265

## RESULT 12

B65145

rhesB protein precursor - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004

C:Accession: B65145; S47701; B36902; A30092; I54935

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C

.A.; Rose, D.J.; Mau, B.; Shaoh, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: B65145

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1411 &lt;BLAT&gt;

A:CROSS-references: UNIPROT:P16917; UNIPROT:P16918; GB:AE000424; GB:U00096; NID:G236723

A:Experimental source: strain K-12, substrain MG1655

R:Plunkett, G.

submitted to the EMBL Data Library, March 1994

A:Reference number: S47666  
A:Accession: S47701  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1129, 'Q', 1131-1411 <PLU>  
A:Cross-references: EMBL:U00039; NID:G466582; PIDN:AAB18457.1; PID:G466618  
A:Experimental source: strain K-12, substrain MG1655  
R:Zhao, S.; Sandt, C.H.; Feulner, G.; Vlazny, D.A.; Gray, J.A.; Hill, C.W.  
J. Bacteriol. 175, 2799-2808, 1993  
A:Title: Rhs elements of Escherichia coli K-12: complex composites of shared and unique  
A:Reference number: A36902; MUID:93259920; PMID:8387990  
A:Accession: B36902  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1397-1411 <ZHA>  
A>Note: sequence extracted from NCBI backbone (NCBIN:132073, NCBI:P:132075)  
R:Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.  
J. Bacteriol. 171, 636-642, 1989  
A:Title: the gene family of Escherichia coli K-12.  
A:Reference number: A91901; MUID:89123133; PMID:2644231  
A:Accession: A30092  
A:Molecule type: DNA  
A:Residues: 1-100 <SAD>  
R:Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;  
J. Bacteriol. 172, 446-456, 1990  
A:Title: Structure of the rhaA locus from Escherichia coli K-12 and comparison of rhaA w  
A:Reference number: 154935; MUID:90094253; PMID:2403547  
A:Accession: 154935  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1221-1411 <RBS>  
A:Cross-references: GB:M29717; NID:G147622; PID:G147623  
C:Comment: the rbs core consist of two distinct parts: a large N-terminal core that is c  
C:Genetics:  
A:Gene: rbsB  
A:Map position: 77 min  
C:Superfamily: rbsF protein  
C:Keywords: transmembrane protein  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-1411/Product: rbsB protein #status predicted <MAT>  
F:28-55/Domain: transmembrane #status predicted <TM>

Query Match 3.8%; Score 340.5; DB 2; Length 1411;  
Best Local Similarity 20.4%; Pred. No. 2.6e-10;  
Matches 287; Conservative 172; Mismatches 459; Indels 490; Gaps 70;

Qy	418	KTEW-ROADNSYRSEITETTFDESGNPLTKVI-----KDKTKIIS-----PS-----	461
Db	218	RTQTPHREAAAGFSGEITGVT-DGAWRHFLVLTQAQRAEAEARQAISGGTEPSAFPDT	276
Qy	462	-----THWEYPPAGEVDNCPPEP---YGPT-----RFLK 488	
Db	277	LPGYTEYGRDNGIRLSAVWLTHDPEP-----ENLPAAPLVRYGWTTPRGELAVVYDRSGK	331
Qy	489	KIIQPPYDSEFK-----DDPEKFIQYRSLGSGSHVTLKIEERHYSATQLLNSTL	539
Db	332	QVRSFYDDKYGRVVAHHTGRPE--IRYRD---SDGRVTEQLNPAGLSVT-----	379
Qy	540	FQYNTDKSELGRLKQTEC--TKGNGKTYSVVHKFTYTKDDTLQOHSITTHDNFTIH	597
Db	380	QYQEKDRITITDLSLRREVLHTQGEAG-LKRVVKK-----EHADGSVT 421	
Qy	598	RSQVRSRYTGRFSDTDKDIQTQMSYDKLGRLLRLTNSGTPYANTLTIDYELNQLDD	657
Db	422	QSQFDA--VGRURAQTDAGRTTEYSPDVVTGLITRIT---TPDGRASAFYNNHNO---	473
Qy	658	NRPPVITITDNGNQLRNEFDGAGRHSQCCLKSD-----GDGKFYTIHTQYDQGRH	712
Db	474	-----LTSATGPDGLREYDELGRLIQETAPDGDITRYRDNPHSDLPATEDATGSR	528
Qy	713	HTSYSDYLITNGRQOTDPDKVHLSKSYDNWNGQANTHWSYGVSEKITVDPI-TLTATK	771
Db	529	KMTWTSRY---QQLLSFTDCSGYVTRYDHRFGQMTAVHREGLSQRAYDSRGQLIAVK	585

Qy	772	QLOSNNVOTGKEVTTYTPSQOPIQITLDEAGHLQ-----SCHTLTRDGWDRVRKE	824
Db	586	DTQCHETRYE-----YNIAGDLTAVIAPDGSRNGTQYDANGKAVRT	626
Qy	825	TDAGQCTIYYDNNRVIOITLPGTIVNRKVAPFSTDLITDIRVNGISLGQOTF--D	882
Db	627	TQG-GLTRSMEXDAAGRVIRLTSENGS-----HTTFRYDVLDRLIQETGDFGRQRYHHD	680
Qy	883	GLSRLTOSQDGRWATYISAGNDOCPSTVITPDGQPIHYQYQPELDDAVLOVASNEITQ	942
Db	681	LTKGLIRSEDEGLV-----THWYD-EADRLTHRTVKGETAE	716
Qy	943	QFSYNPTVGAL--LKAVAEGOSLTPYY--PSGRL--KMNINDMKMSYLM-----	988
Db	717	RWQYDE-RGWLTDISHISEGHRVA-VHYRYDEKRLTGERQTVHHPQTEALLMQHETRA	774
Qy	989	-TLRLENGYDTLTGTTQIKISRDTHGRVTOIKOSSIKITLYNDDLNHRHISQVTDLATGH	1047
Db	775	YNAQGLAN--RCIPDSLPVAVWLTYG-----SGYLAKMKLGD-----	809
Qy	1048	MLTTTVEF--DGLNREIGRKLCDSSGHTLDIQOSWLKTQQLANRIVKLVGLQTEQVSY	1105
Db	810	--TFLVEYTRDRRLHRETIR-----SFGRYELTAYTPAGLOSO--HLNLSLS--DRDYTW	859
Qy	1106	DSRNRLNOYKCDGAECPDKYGHISIVTQNFYDIYGNITACHTTFADGTEDHATFKFANP	1165
Db	860	NDNGELIR-----ISSPRQ-----TRSYSYSTTGLTGVHTTAA-----NLDIRIPYA	902
Qy	1166	TDPC--QLTEVHHTHPD-----MPDN-----IRLKYDKAGRVINITD-----NH	1202
Db	903	TDPAGNRLPD-PELHPDSTLSMWDNRRIARDAHVLYRYDRHRLTETDLIPEGVIRTTDD	961
Qy	1203	GNTENFYDTLGRLONGQ-----GSVYGYDPLNRLYSOKTDTLDCEL-----	1244
Db	962	ERTHYHYDSQHRVLVHYTRTQYEBPLVESRYLYDPLGRRVAKRVWRERDUTGWSLSRK	1021
Qy	1245	---YY---RETMLVNE-----VRNGEMIRLLR---TGETIIAQORA-----	1276
Db	1022	POVTWYGDGDLRTTIQNDRSRIQTIYQPGSFTELRIVETATGELAKTQRSLADALQOS	1081
Qy	1277	-----SKULLGTDSQOSVILTSQKNSQEA-----	1303
Db	1082	GGEDGGSVFPPLVQMLDRLESEILA---DRVSESRRLASCLTVEQMKQNDPVYT	1138
Qy	1304	-----YSAYGKHKSTAND---ASILGYNGERADP	1329
Db	1139	PARKHLYHCDHRGLPLALISTEGATWCAEYDEWGNLLNEENPHQLQQLIRLPGQQYDE	1198
Qy	1330	VSGVTHLNGYRSYDPTLMRFPHPTDLSLSPFG-AGGINPYSCYCLGDPINRSDPSGHLWSQA	1388
Db	1199	ESGLYY--NRHRYDPLQGRYITQD---PIGLKGMNLYGYQL-NPISIDIDPLGLSMWED	1252
Qy	1389	WTGIMGAGLLLTITATGGMAIAAAGGIAAIAISTTALAFGALSVTSDITSIVSGALE	1448
Db	1253	-----AKSG---ACTNGLCGTSLA-----MIGPKDFSDIDSTAY	1283
Qy	1449	DASPKASILLGWSMGMAAGLAESAIGKGTKLATHLGAFAEDGENALLKSTSSSRKWK	1508
Db	1284	DALNKINS-----QSICEDKEFA-----	1301
Qy	1509	GVTRSLDREIVRNEBEGQVVKDHSRGYTDNFMKGEGQ-----AILVHDKDGLFYHTEG	1561
Db	1302	-----GLICKNSGRYFSTAPNRGKRGSGYFPNPGPCNGTEKVSAYHTHG	1346
Qy	1562	NKNGK--GPYTRHTPEQLVDYLDKNNI	1587
Db	1347	ADSHGEYWEIFSGKDEKIVK-SKDNNI	1373

RESULT 13  
H64780  
rhaD protein precursor - Escherichia coli (strain K-12)



C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: E85509  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: E85509  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1404 <STO>  
A:Cross-references: UNIPROT:O8XED9; GB:AB005174; NID:g12512977; PIDN:AAG54537.1; GSPDB:C  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z0268  
C:Superfamily: rhes protein

Query Match 3.8%; Score 336.5; DB 2; Length 1404;  
Best Local Similarity 20.0%; Pred. No. 4.3e-10;  
Matches 252; Conservative 130; Mismatches 373; Indels 505; Gaps 58;

QY 376 YHL-LTSECKQONGVIQTETAYAIIGHNFDSPQSQFQLPKT-KTETWRSDNSYRSEI 433  
DB 245 FHLVLTQQAERVKORATSLSPAGER--SASSSLVFPDTPAGTEYGADNGIRLEA 302  
QY 434 TETTFDES-----GNPLTKVKKTKQKIISPSHWEYYPAGEV---DNCPEPYGF 483  
DB 303 VWLTHDPAYDELPAPLARYT-----YTASGELRAYDRSGTVRGPF 345  
QY 484 TRFVKKIIQTPYDSF-----FKDDPEKFIQVYSLIGSQSHVTLKIBERHYSATQL 534  
DB 346 A-----YDAESHAGRMVAHVAGRPES--RYRYDDTG-----RVTEL 379  
QY 535 LN-----STLFQYNTDKSELGRLLKQTEC--TKGENKTSYVHVHFTYTKQDDTLQOQSHI 588  
DB 380 VNPEGLDYRFYQDQVRVITDLSNRREVLYTEGEGG-LKRVVVK-----422  
QY 589 TTHDNFTIHRSVRSRYTGRFLFSDDTDKDIVTQMSYDKLGRLLTKTLNSGTPYANTLYD 648  
DB 423 -EHADGSLTRSE-----YDEAGRLKAGT-----444  
QY 649 YELNNLQDNRPPFVITTDVNGNQLRNEFPDAGRHVSQCLKSDG-----DGKPYT 700  
DB 445 -----DAAGRRTEYSLHMASGAVTAVTGPDGR--T 472  
QY 701 IHTQYDQGRHHSTYSYDLTNGRQQTDPKXVHLSMSKSYDNWQCIANTHWSGVSEKI 760  
DB 473 VR-YGYNQRQVTSVTYTPDGLRSRREYDEKGLAAETSRSGE-----TTRYSDYD-----520  
QY 761 TVDPITLTATQLQSNVNVOTGKRVTTYTPSQQPIQITLTFDEAGHLQSCHTLTRDGWDR 820  
DB 521 --DP-----ASELPTGQDAGSTKQM-----AWSR 544  
QY 821 ---VRKETDAIGQCTIYQYDNNYRVIQITLDPGTIVNRKYAPFSTDLITDIRVNGISLG 877  
DB 545 YGQLLTFTDCGYTRYEDRYGQCIQAVHREEGISTYSSYNP-----RG 588  
QY 878 QQTDEGLSRLTQSDGGRVWATYTSAGNDQCPSTVITPDGQFIHYQYQPELDDA---VLQ 934  
DB 589 Q-----LVSKQAQGRRETRYEYSAAGDL--TAIVAPDGRSRSEIYQ-----DAMGKAVS 634  
QY 935 VASNEITQCFSYNPVNTGALLKAVAGQSLTPIYPSGRKMKENIDMKMSYLTWLRGLE 994  
DB 635 TTQGLTRSMGYDAAGRTVLTNENGSQSTRYDPVDELTEQRFQDGTQRY-----686  
QY 995 NGYTDLTGTI-----QKISRDTHGRVTQIKDS--1022  
DB 687 --HYDLATGKLQSEDEGLVTLWHYDASDRITHRTVNGDPAPQWQVDEHGWLTLTSHSTSEG 744  
QY 1023 --IKTLNVDLNRHIGSQ--VTDLATGHML-----1049  
DB 745 HRVSVHYGDDKGRLTGSRQTVENPTEGMLWEHTGHAYSEQGLATKQEPDGLPPVEWL 804

QY 1050 -----TTTVEF--DGLNREIGRKLCDSSGHTLIDIQOSWLKLTQQLANRIVKL 1093  
DB 805 TYGSGYLAKMKLGGTPLVYEMRDLRHETARSF--GGEAYELATAWNTSQLSRHLNL 861  
QY 1094 NGVLQRTQEQSYSDSRNRLNQYKDGACPTDKYGHISIVTQNFYDIYGNITACHTTTFADG 1153  
DB 862 P---QLDRDYDWDNGQL--IRISGPQ-----ESREYRYSDTGLTGVHTTAANL 906  
QY 1154 TED--HATPKFANP-TDPCQLTEVHHTHPD-----MPDNIRL-----KYDKAGRVIN 1197  
DB 907 DIDIPYATDPAGNRLPOP-----ELHPDSTLTAWPDN-RIABDAHYVYRYDEYGLAE 958  
QY 1198 ITD-----NHGNTENFTYDTLGLRL-----QNGOGSV---YGYDPL-----1229  
DB 959 KTDRIPGVIRMHDERTHHYDSQHRLVFETRLQHGEPQVESRYLYDPLGRGTGKRVWR 1018  
QY 1230 -----NRLVSQKTDLDCELYYR-----ETMLVNEVRNGEMI 1261  
DB 1019 RERDLTGWMSLSRKPEETWYGMGDRLTTVQTQTRIQTVYOPGSGFTPLLRIETENGEOA 1078  
QY 1262 RLLR-----TGETITAQ-----QRASKVLLTGTDSQOS-----1289  
DB 1079 KARHSLAEVLQEDPTGVTLPAAELAVMLGRLERLRQGSVSEESQOWLAQCGLTAEQMAAA 1138  
QY 1290 -----VILTSKQNLQSEAYSAYGK--HKSTANDASILGYN 1323  
DB 1139 LEAYIYIPERKHLHYCHDRGLPLALISPEGETAQQGEYDEWGNLLGETSAQHLQOQSLRLP 1198  
QY 1324 GERADPVGSVTHLQNGVRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPSG 1382  
DB 1199 GQYDEESGLYNNRY--YDPLQGRYITQD---PIGLEGGWNLQYVPL-NPIEHIDPLG 1252

RESULT 15  
C65159  
rhesA protein precursor - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004  
C:Accession: C65159; S47814; I69402  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: C65159  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1377 <BLAT>  
A:Cross-references: UNIPROT:P16916; GB:AB000437; GB:U00096; NID:g2367249; PIDN:AAC76617.  
A:Experimental source: strain K-12, substrain MG1655  
R:Plunkett, G.  
submitted to the EMBL Data Library, March 1994  
A:Reference number: S47666  
A:Accession: S47814  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-864, 'M', 866-1377 <PLU>  
A:Cross-references: EMBL:U00039; NID:g466582; PIDN:AAB18570.1; PID:g466731  
A:Experimental source: strain K-12, substrain MG1655  
R:Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;  
J. Bacteriol. 172, 446-456, 1990  
A:Title: Structure of the rhesA locus from Escherichia coli K-12 and comparison of rhesA w  
A:Reference number: I54935; MUID:90094253; PMID:2403547  
A:Accession: I69402  
A:Status: preliminary; translated from GB/EMBL/DBDJB  
A:Molecule type: DNA  
A:Residues: 1-864, 'M', 866-1377 <RES>  
A:Cross-references: GB:M29716; NID:g147613; PID:g147614  
A:Experimental source: strain K-12  
C:Comment: the rhes core consist of two distinct parts: a large N-terminal core that is c  
A:Gene: rhesA

C: Superfamily: rhsp protein

C: keyword: transmembrane protein  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-1377/Product: rhd protein #status predicted <MAT>  
F:28-55/Domain: transmembrane #status predicted <TMM>

Query Match 3.7%; Score 331.5; DB 2; Length 1377;

Best Local Similarity 20.8%; Pred. No. 7.8e-10;  
Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;

QY 211 LFPQKEGY-RTELRLNQLNSIHNFSLGNENPLTWSFGYTPIGKNGILGQWITSMTAP 269  
DB 126 LFPGE-DGVSRESLWLR--GGVAKLDEGRLAALWQ-----160  
QY 270 GGLKETVNS-----NNNQ-----HHFPOSANL---PVLPPVTLMKQVFCAGQ- 310  
DB 161 -ALPEELRLSPHYLATNSPQPGWLLGWCERPEADVLPAPLPPYRLVTGLVDRFGRT 219  
QY 311 PAIQABYSTSYNGGSGNGIWNKKLDNLYGLMTEYNGSTESRYKDKKHQDIQVRIE 370  
DB 220 QTFRHAEAGEFSGEITGVTDGAWRH-----FRLVL-----TTQAGR-----255  
QY 371 RTYNNVHLTSSCKQONGYIQTETAYYALIGHNFDSPSQFOLPKT---KTETWRSADN 427  
DB 256 -----AEARQQ-----AISG-----GTEFSAP--POTLPGYTEYGR--DN 287  
QY 428 SYRSEITETTFDESIGNPLTKVIKDKTKKIIISPTHWEYYPAGEVDNCPPEP---YGFT 484  
DB 288 GIRLSAVLTHDPE-----YP-----ENLPAAPLVRYGWT 317  
QY 485 -----RFVKIIQTPYSEPK-----DDPEKFIQRYSLIGSQSHVTLKIE 525  
DB 318 PRGELAVVYDRSGQVRSFTYDDKVRGNVAHRHTGRPE--IRYRYD---SDGRVTEQLN 372  
QY 526 ERHSATQLLNSTLFOYNTKSELGRLAKQTC--TKGNGKTSYVVKHFTYTKQDDTLQ 583  
DB 373 PAGLSY-----VOYEKDRITITDSLRREVLHTQGEAG-LKRVVK-----413  
QY 584 QSHSITHTDNFTIHRSQVRSRYTGRLFSDTDKDIVTQMSYDKLGRLLTRLNSGTTPYAN 643  
DB 414 -----EHADGVSVTQSFDA--VGLRAQTDAAGRTEYSVDVVTGLITRIT---TPDGR 462  
QY 644 TITYDYELNQLQDNRPFPVITTDVNGNQLNEFDGAGRHSQCCLKSD-----GDGKF 698  
DB 463 ASAFYYNNHQ-----LTSATGPDGLERREYDELGRLIQETAPDGDITRYRYDNPH 514  
QY 699 YTIHTQQYDEQGRHHTSYSDYLTNGRQOTDPDKVHLSMSKSYDNWQJANTHWSYVSE 758  
DB 515 SDLPCATEDATGSRKTMWTSRY---GQLLSFTDCSGYVTRYDHRFGQMTAVHREGLSQ 571  
QY 759 KITVDPI-TLTATKQLQSNNSNVQTKGKVTVTTPSQOPIQITLFDAGHLQ-----SC 810  
DB 572 YRAYSRGQIIAVKDTQGHETRYE-----YNIAGDLTAVIAPDGR 612  
QY 811 HTLRDGDWRVRKETDAIGCCTIYQVDYNNRVIQITLPDGTIVNRKYAPFSTDTLITDIR 870  
DB 613 NGTQYDAMGKAVRTIQG-GLTRSMEYDAAGRIVRLTSENGS-----HTTFRYDVLDRLIQ 666  
QY 871 VNGISLQCOQTF--DGLSRLTQSQDGRVWAYTYSAGNDQCPSTVITPDQOFIHYQOPEL 928  
DB 667 ETGFDGRTQRYHHDLTGKLIRSEDEGLV-----THHYD-EA 702  
QY 929 DDVAVLQVASNETIQSPYNPVTGAL--LKAVAGQSITPIYY--PSGRL--KMNINDM 981  
DB 703 DLRLTHRTVKGETAERQWYDE-RGWLTDISHISEGHRVA-VHYRYDEKGRLTGERTQTVHP 760  
QY 982 KKMYSILW-----TLRGLNGYDITLGTIQKISRDTGHRVTQIKDSSIKTTLNYYDDL 1033  
DB 761 QTEALLWQHETRHAYNAOGLAN--RCIPDSLPAVEWLTYG-----SGYLAGMKLGD-- 809  
QY 1034 RHIGSQVTLATGHMLTTTTFEP--DGLNREIGRKLCDSSGHTLDTQQSWLKTQQLANRIV 1091  
DB 810 -----TPLVEYTRDLRLHRETLR-----SFGRYELTAYTAPAGLOQSQ-- 846

Search completed: February 9, 2005, 18:18:45

Job time : 78 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2005, 18:14:07 ; Search time 182 Seconds  
(without alignments)  
2993.957 Million cell updates/sec

Title: US-09-889-874A-23  
Perfect score: 8879  
Sequence: 1 VTIKFLFRITMSDNEP.....PKIILGRTEKTVKPTFRP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1373511 seqs, 325702437 residues

Total number of hits satisfying chosen parameters: 1373511

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications\_AA.\*
- 1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/prodata/1/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 13: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB.pep.\*
  - 17: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB.pep.\*
  - 18: /cgn2\_6/prodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
  - 19: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
  - 20: /cgn2\_6/prodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	932.5	10.5	982	15	US-10-365-742-106
2	517.5	5.8	1317	15	US-10-369-493-11243
3	420.5	4.7	2364	14	US-10-156-761-7834
4	395	4.4	820	14	US-10-156-761-7990
5	392	4.4	1250	14	US-10-156-761-7572
6	391.5	4.4	2386	14	US-10-156-761-7751
7	386	4.3	2234	15	US-10-282-122A-46565
8	361	4.1	1510	15	US-10-418-861B-55
9	353.5	4.0	843	15	US-10-282-122A-76214
10	350	3.9	1515	15	US-10-282-122A-47600
11	342	3.9	1397	15	US-10-282-122A-43059
12	340.5	3.8	1411	15	US-10-282-122A-43060
13	339	3.8	1426	9	US-09-912-020-340

14	339	3.8	1426	15	US-10-282-122A-42617	Sequence 42617, A
15	331.5	3.7	1377	9	US-09-815-242-10384	Sequence 10384, A
16	331.5	3.7	1377	14	US-10-287-274-467	Sequence 467, App
17	331.5	3.7	1377	15	US-10-282-122A-42731	Sequence 42731, A
18	328	3.7	1572	15	US-10-282-122A-69415	Sequence 69415, A
19	322	3.6	1385	15	US-10-282-122A-68242	Sequence 68242, A
20	317.5	3.6	1365	15	US-10-282-122A-44558	Sequence 44558, A
21	314	3.5	1395	15	US-10-282-122A-50586	Sequence 50586, A
22	307.5	3.5	932	15	US-10-282-122A-44617	Sequence 44617, A
23	306.5	3.5	1530	15	US-10-282-122A-68175	Sequence 68175, A
24	298.5	3.4	1531	15	US-10-282-122A-49308	Sequence 49308, A
25	297	3.3	1512	15	US-10-282-122A-78386	Sequence 78386, A
26	291.5	3.3	1253	15	US-10-282-122A-68914	Sequence 68914, A
27	291	3.3	2346	15	US-10-072-012-491	Sequence 491, Appl
28	290.5	3.3	2802	9	US-09-808-602-81	Sequence 81, Appl
29	290.5	3.3	2802	10	US-09-800-198-69	Sequence 69, Appl
30	290.5	3.3	2802	15	US-10-072-012-489	Sequence 489, Appl
31	289.5	3.3	1364	15	US-10-282-122A-75309	Sequence 75309, A
32	285	3.2	2613	15	US-10-038-854-42	Sequence 42, Appl
33	285	3.2	2628	15	US-10-038-854-40	Sequence 40, Appl
34	285	3.2	2721	15	US-10-038-854-36	Sequence 36, Appl
35	285	3.2	2725	15	US-10-038-854-36	Sequence 36, Appl
36	284.5	3.2	2715	15	US-10-042-865-52	Sequence 52, Appl
37	284.5	3.2	2715	15	US-10-023-020-51	Sequence 51, Appl
38	280	3.2	1317	9	US-09-815-242-5118	Sequence 5118, Ap
39	280	3.2	1317	15	US-10-282-122A-43495	Sequence 43495, A
40	279.5	3.1	944	15	US-10-282-122A-47806	Sequence 47806, A
41	278	3.1	985	15	US-10-282-122A-43061	Sequence 43061, A
42	276.5	3.1	1438	15	US-10-282-122A-78360	Sequence 78360, A
43	273.5	3.1	1688	15	US-10-144-194A-113	Sequence 113, Appl
44	273.5	3.1	1737	9	US-09-808-602-83	Sequence 83, Appl
45	273.5	3.1	1737	10	US-09-800-198-71	Sequence 71, Appl

ALIGNMENTS

RESULT 1

US-10-365-742-106  
; Sequence 106, Application US/10365742  
; Publication No. US20030204868A1  
; GENERAL INFORMATION:  
; APPLICANT: Collmer, Alan  
; APPLICANT: Alfano, James R.  
; APPLICANT: Cartinhour, Samuel W.  
; APPLICANT: Schneider, David J.  
; APPLICANT: Tang, Xiaoyan  
; TITLE OF INVENTION: PSEUDOMONAS AVR AND HOP PROTEINS, THEIR ENCODING  
; FILE REFERENCE: 19603/4112  
; CURRENT APPLICATION NUMBER: US/10/365,742  
; CURRENT FILING DATE: 2002-02-12  
; PRIOR APPLICATION NUMBER: 60/356,408  
; PRIOR FILING DATE: 2002-02-12  
; PRIOR APPLICATION NUMBER: 60/380,185  
; PRIOR FILING DATE: 2002-05-10  
; NUMBER OF SEQ ID NOS: 209  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 106  
; LENGTH: 982  
; TYPE: PRT  
; ORGANISM: Pseudomonas syringae pv. tomato DC3000  
US-10-365-742-106

Query Match 10.5%; Score 932.5; DB 15; Length 982;  
Best Local Similarity 29.7%; Pred. No. 5.7e-56;  
Matches 285; Conservative 142; Mismatches 357; Indels 175; Gaps 33;  
QY 642 ANTLFYDELNNLQDDNRPFFVITTDVNGNQLRNEFDGAGRHSVQSCLXSDGDGKFYTI 701  
DB 31 SSTLLGRYVD-----DWNQR--CCTTTDDN-VQTYEYSDPIGSDVHK-----GPIQKT 74  
QY 702 HTQQYDEQGRHHTSTVSYDLTNGRQQTDPDKVHLSMSKSYDNWQIANTHWSYGVSEKIT 761

```
Db 75 WKQSGDPEGR-----ISGRSET-----WLNL-----FGKPDRI 103
Qy 762 VDPITLTATKQLOSNNSNVQTKVTVTPSOPIQITLFDAGHQSCHTLTRGMDRV 821
Db 104 -----TLTAKTGRSHMSRNRNLT-----TEQEL-----SRQTFYDGLGRC 144
Qy 822 RKETDAIGCTIYQYDNTNRVFIQITLPDGTIVNRKYAPFSTDLITDIRV---NGIS--- 875
Db 145 TEQRDALQOSTLFSYDNRSMVSSTLADGSVINRSYAPQSSELATMLEVHVHONGTTRTV 204
Qy 876 LGQTFDGLSLRUTSQODGRVWAYYSAGNDQCPSTVITPDQOFIHYQYQPELDVAIQV 935
Db 205 AGTKQFDGLERVTQTKGDRVQFNYDAGEMQ-PRSRITTAGLDNINFTYTRALTQDIFSS 263
Qy 936 ASNETIQFSPNVPVTCALLKVAEQSLTPIYPGSRLLKMNINDMKMSYLM-----T 989
Db 264 TAPDETAKFYDKTSARLIEATNPQGTTRYVDVHNQLTGETWDLN--LGQAWETHROSS 321
Qy 990 LRGLENGVTDL-----TGTIQKISRDTGRVTOIKDSSIKTTLNVDLNRHIGSQVTDLA 1044
Db 322 LLGRPIKRTDLKGAEGAETRYDYTLGRIFRINQSNLRTIIDYDLGOLCKVATEDLQ 381
Qy 1045 TGHMLTTTVEPGLNREIKRLCDSSGHTLDIQOQSWLKTQOLANRIVKLVGLQRTQYS 1104
Db 382 AGTGVIIDMEYDQGOEILRTQTASQAALTLQTWAVDGLKTRDLQAGSPLHETFS 441
Qy 1105 YDSRNLNQKCDGACETPKYGHISVTQNFYDIYGNITACHTTFADGTEDHAFKAN 1164
Db 442 YDPRGLTLVNLVGLSRLDELOREMTQIFSDLDNLTLCQTRFTDGTSEAAFKYS 501
Qy 1165 P-----TOPCOLTEVHHTPMDPNIRLKYDKAGRVINITDNGHNTENFTYDTLGR-- 1216
Db 502 PGDDKHKRCQLLSATYTPRTPPTPSYDANGN--QLKDEGN--SLHYDSQSLLOV 557
Qy 1217 -QNGQG--SVGYDPLNRLVSQKTDPLDCEL--YYRETMVNEVRNGEMIRLLRTGETII 1271
Db 558 AETGAPISOYRDYGHNLVATR-DONESEILRFYEGHQLSSTVQEDQRTQYHLGEOPL 616
Qy 1272 AQO---RASKVLLTGTDSQSVILISDKQNLQSEAYSAYGHKSSTANDASILYNGERAD 1328
Db 617 GOQIVDDAEQTLTLLTDANQSVMGFQOQLRKAVYSAYGERHSBEALLSTAGFNGVRE 676
Qy 1329 PVSQVTHLNGYRSYDPTLMRPHTPDSLPFGAGINPVSYCLGDPINRSDPSGH----- 1383
Db 677 AANGHYLLNGYRANPLMRHSHDFDLSPPFAGGVNPPYICLGNPIALRDPFGHDASQ 736
Qy 1384 -----LSWQ-----AWTGIGMGIA-----GLLLTIATGMAIAAAGGI----- 1416
Db 737 TGRLLRPDEGALPMQGGGDMGWGVGVVTVLGVAAATLGTATPVTGPVTVLGI 796
Qy 1417 -----AAAIATSTTALAFG-----ALSVTSITISVALGEDADSPKASSILGWVS----- 1462
Db 797 SMTASAAAASVTSVTCALITVGTALTAASTTANTVAIVNN-----DQTAGEVSGMGLIAAV 851
Qy 1463 -----MGWGAAGLAESAIGKTKLA-----THLGAFAEGBENALKSTSESSR 1505
Db 852 PVCLVGFAGAVARAAVAAAANAAGTIGVRSVRIG-LAAAGARTISSAASSAR 909
```

## RESULT 2

```
US-10-369-493-11243
; Sequence 11243, Application US/10369493
; Publication NO. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
```

```
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11243
; LENGTH: 1317
; TYPE: PRT
; ORGANISM: Methanosarcina mazei
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1317)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-11243
```

Query Match 5.8%; Score 517.5; DB 15; Length 1317;

Best Local Similarity 21.2%; Pred.No. 3.1e-26; Matches 334; Conservative 194; Mismatches 505; Indels 541; Gaps 72;

```
Qy 39 TGLVNIQITLGHVGNLGLPTLPLTL--SYSPLNKTDIGFGIG--FNFGLSVYDRKNSL 94
Db 56 TGSFYQYQDLSIPRG-----LPLTVSRSYNSMDNRSLGFGSGWTFYNNMLTVDNNGN 110
Qy 95 LSLSTGENYK---VIETDKTVLQOKKLDNLRPEKDKENCYRIIHKSGDIEVLTFGNFN 151
Db 111 VTVLGGDGHDTYILNPDGTYSRPLSVFDDL-----IKNSDETTLTKKQDT 157
Qy 152 AFDLVKPKILN---PAGHAIYIDWNFEATOPRLNRIYDDLGDHDIPLNLEYQGLIKTI 208
Db 158 KYNFSSEKLVNIVDKNGNQI---NFTYGEQLTKV--TDASGREL--ILAYDHNG----- 206
Qy 209 LTLPGQEGYRTELRLNRLQNLNHNPLSGNENPLTWSFGYTPIG-----KNGILGO-- 261
Db 207 -----HIISITDPMGRVMSYSDDOGNLIQCKNPIGGKLS 241
Qy 262 -----WITSMTAPGG--LKETVYNSNNQGHFPOSANLPVLPVYVLMKQVPGAGOP 311
Db 242 YTYDENHMTSITDPRGNHPKNTY---DEKGVHSQS----- 276
Qy 312 AIQAEYSYTHNYGGSGNGIWNKLDNLYGLMTEYNYGSTESRYKDKEGHDQIVRIER 371
Db 277 -----NSLNATY-----TFNYDS-ENRK----- 293
Qy 372 TWNNYHLLTSECKQONGYQIOTETAYYAIIGHNFDSPSQFOLPKTKTETWRSADNSYRS 431
Db 294 -----TTETD-----PFGKNTYSFDEHFW-- 313
Qy 432 EITET-----TFDESGNPLTKVIKDKTKQKIISPSTHWEYPPPAGEVDNCPPEPVG 482
Db 314 ELNETNQLGYTISYAYDENGNRISVTNENSKTKLA-----YDANGNIK-TTNPLG 364
Qy 483 FTRPVKKLIOTPYDSEPKDDPEKFIQYRYSLIGSQSHVTLKTEERHYSATQLNLS----- 537
Db 365 YSK-----SMTYDS-----KNNLISQTDRLRHKTSFEDYDNNLSIKSIDALG 406
Qy 538 --TLFOYNTDKSELGRLLKQTECKT-----GENGK--TYSVVHKPT 574
Db 407 HETVPSYD--KYGOVIGETDSNKKTATFSYNNNGDQIITTDANGKTSAFYDVTGRTV 462
Qy 575 -----YTKQDDTLQOSSHISITTHDFTIHRSQVRSRYTGRL-----FSDTDTKDI 618
Db 463 TKTDAGKNRYTFQYDALONLLSITD-----PMGQTTSTNTMLLEIKLVQLMKVDXQSI 516
Qy 619 VTQMSYDKLGRLLITFLNSGT---PYA-----NTLYDYELNLLQDNRPPFVIT 665
Db 517 LTLTXISLXKXRMOWVEXLGTNNMPLAIWPPRQTKGHKTSYDYDPLNRQ-----VS 568
Qy 666 TTDVNGNQLRNEFDGAGRHS-----QCLKDSGDGKFYTIHTQOYDE 708
Db 569 VTNALGKTRKRYDAIGNKISITNAYGKSTRYSYNSLNQLVKVTNAMGK---VVRYNYDA 625
Qy 709 QGRHHTSYSD-----YLTNGRQQTDPDKVHLMSKMSYDNWNGQIANTHSYGVSEKIT 761
Db 709 QGRHHTSYSD-----YLTNGRQQTDPDKVHLMSKMSYDNWNGQIANTHSYGVSEKIT 761
```

Db 626 VGNLSTTDENGKINYGDSLNRQSVTDALRKTTRNKYDAVGNKISITNAYGKSTRYS 685  
QY 762 VDPITLTATKQLQSNVNTGKEVTTYTPSOQPIQITLFDAGHLQSCHTLTRDGDWRV 821  
Db 686 Y-----NSLN-----QLVKVTDAMGVVYNYDAVGNLIS----- 715  
QY 822 RKETAIGOCTIYQVNDYNRVITLTPDGTIVNRKYAPFSTDTLITDIRVNGISLGOQTF 881  
Db 716 --TTDANGRKTYGDSLNRQSVITNALGKTRNKYDAVGNKISSTDANWR---LTKYSY 770  
QY 882 DGLSRLTQSDG-GRVWATYGA-GNDQCPSTVITPD--GQFTHYQY-----QPELDDAV 932  
Db 771 DSNRLVKVTDAMGVVYTYDAVN-----LISTDAKHKTDYGYDSLDRQVSTIDPL 825  
QY 933 LQVAGNEITQOFSYNPVTGALLKAVAQSGSLTPPIYPSGRLXWENIMDKMSYLTWLRG 992  
Db 826 GRTARK-----YDAVGNKISSTDEGKTSYGYDVLNRLTKVSPDDQKVSY----- 873  
QY 993 LENGTYDLTGTTQKISRDTGRTVQIKDSSIKTLLNYDDLNRHI-----GSOV--TDLA 1044  
Db 874 -----NYDAVGNRLTKWDSHGTTAYKYDKLNRLLSVLNPDGQKVSYTNK 918  
QY 1045 TGHML-----TTTVEFDGLNREIGRKLCDSSGHTLD---IQOSWLKTQOLANRIVKL 1093  
Db 919 VGNRVKMTYPDGKTSYSYDAVNRLG--VIDSDGHTSYSAKGNLTKMTNTPGV--- 973  
QY 1094 NGVLQTEQYSDSRNRLNQKCDGAECPDKYGHISIVTQNFYDIYGNITACHTTTFADG 1153  
Db 974 -----KTE-YSYDKANRL-----VELINKNTQVSSYKYTLDAAGNRLKVDQLAEG 1020  
QY 1154 TED-----HATFPAFNPDPQCLTEVHHTHPMDPN--IRLKVDKAGRVNIT- 1199  
Db 1021 VESGSELKESQLLTYYGY---DKLYRLTKV-----DYPSNKTYSYKYDSMGNRISMTT 1072  
QY 1200 ---DNHGNTEFNFTYDLGLRQNGQSVGYVDPLNRLVSOKTDTLDCELY-YRETLVNEVR 1256  
Db 1073 NVDGISTISYKDAADQLQSGNISYSYKGNLKKRVNTPQFMSYSIDE---ANRLK 1129  
QY 1257 N-----GEMIRLLRT--GETIIAQORASKVLLTGTDSQOSVILT 1293  
Db 1130 NVSEFVSNTPKSSYNFEYDGGNRIIKTINGE---NAQSTKYVLDINSALQVLT 1185  
QY 1294 SKQNL-----SOEA-----YSAGKHKSTANDASIL----- 1320  
Db 1186 SDTKNTTCYTGDTLISMTNSENAEEYHYHDLGSLVRSLSDSKGIKNTLYDAFGQVOK 1245  
QY 1321 -----GYNGERADPSGVTHLNGYRSYDPTLMRPHTPDLSL--PFGAGGINPYSY 1369  
Db 1246 EIGTVDNFRFTGEQMDDETGLIYLARY--YDPSVGRFITKDVIKGRVTTQSNIRVYV 1303  
QY 1370 CLGDPINRSDPSGH 1383  
Db 1304 TTNNPVLVDLTGY 1317

## RESULT 3

US-10-156-761-7834  
; Sequence 7834, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156, 761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30

Query Match 4.7%; Score 420.5; DB 14; Length 2364;  
Best Local Similarity 21.6%; Pred. No. 1.8e-19;  
Matches 320; Conservative 190; Mismatches 550; Indels 423; Gaps 73;  
QY 134 RIHKSGDIEVLUTGNNAFADLKVPKKLLNPAGHAIYIDWNFEATQPLNRRIYDDLDGHD 193  
Db 702 RXYTGDDIVLVPINFNRDLD-----NKVGS-----ELNF-----PRIKEIHGDL----- 744  
QY 194 IPLLNLVQGLIKTTLTPPGQKGYRTELRLNQLNSIHNFLSNGNENPLTWSFYPTPI 253  
Db 745 -----GATTVSYGFAN-----ACDIDLHPAQAS-----NTQDCYMQ-KWTPE 781  
QY 254 GKNGILGOWI-----TSMTPAGGLKETVNYVS-NNNQGHFFPQSANLPLVLP-- 297  
Db 782 GETDSKTGWFKKFLVTVQVDFVTNQQDGAPEVMTTSYTYEDGAGWHP---TNDPLIKDB 838  
QY 298 -----YVTLMKQV-PGAGQPAIQAEYSYTSYHNYGGGNGIWNKNDLNYGLMTEY 347  
Db 839 DESWTDWRGYQVQVTTGAGAQK-----TKKSWLYRGLSGDRTSKAD----- 881  
QY 348 NYGSTESRYKDKGHDQIVRIERTYNNY-----HLLTSECKQONGYIQTETAYYA 399  
Db 882 -ASATKTVTVDDGCG-----NNYTDSWLSGRILSTSLRDDTG--TSHERTYHK 927  
QY 400 IIGHN---FDSQSPQQLPKTKTETWRSADNSYRSEITETTPDES-----GNPLTKVI 449  
Db 928 YWHDNTAQYDGLDFRREKETTNTYKVSQWRHETVETEDTDEGASTTGLPMR--- 984  
QY 450 KDKTKQKIISPSTHWEYPPAGEVDN-CPPEPYGFTRFVKKIIQTPYDSEFKDDPKFIQ 508  
Db 985 TDDWQSSVS-----DNRC-----TTYGRAYNTD----- 1008  
QY 509 RYSLIGSQSHVTLKIEBHYS-----ATOLLNSTLFPQN----- 543  
Db 1009 -NYDSTGAQRWTVLQDVQKHYSGVCSGSIADSKQDGYASTLYDNATSVDAKPFVDGNPTES 1067  
QY 544 -----TDKSELGRLLKQTE-----CTKGKNGKTSYVVVKFTYTKQDDTLQ 584  
Db 1068 RYTKSGSYRSTWSGYDDAGRVMSSEDSGSHNRTLKYSPANTWPLNGVIVTTPDDCALP 1127  
QY 585 SHSITTHDNFTIHRQVRSRYTGRFLSDTDTKDIVTQMSYDKLGRLL----- 631  
Db 1128 AHTALTSTAWT-----SRFPGKFTSIQDANGNVTKMSLDAAGRLVEVWRPETGSSSPS 1180  
QY 632 -----TFTLNSGTPYANTLYDYELNNLQDDNRPFFVI 664  
Db 1181 MKFSYTIPTSTNSAGVPDAVDGYPHVATHVLOSQTYLSSHAYVDGLGRAR-ETOTPMGN 1239  
QY 665 TTTDVNGNQLRNEFPDAGRHVSQCLKSDGCKGFTYIHTQOYDEQGR-----HH 713  
Db 1240 GVDAAATGNEVPN-----RQVSVTRYDSAGN---VTGTSVFRNQTAGSGPSSAKVED 1290  
QY 714 TSTYSYDYLTN--GROQTPDPDKVHLSMSKSYDNWQGIANTHWSYGVSEKIVTDPITLTATK 771  
Db 1291 LPSYTDVLVDWAGRAITS-----RLQVNGASQDAGRVDITYDGDFTSVKKNVDAADTYD 1346  
QY 772 QLQSNNSNVQTKQEVN-----TYTSPQOPIQITLFDAGHLQSCHTLTRDGDWRVKEFTD 826  
Db 1347 VYQVSKVVEHTGSAITYTAYTAKDELIKIT--DPRGN-DTSYTYDWAGORTATDSD 1403  
QY 827 AIGQCTIYQVNDYNRVITLTPDG-TIVNRKYAPFSTDTLITDIRVNGISLGOQTFDGLS 885  
Db 1404 AGVSST--EYDKNGQIKSITSGGKTVLDYGYDALGRKTA---VRSQDTELAANTWDGLN 1458



APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 7572  
LENGTH: 1250  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-7572

Query Match 4.4%; Score 392; DB 14; Length 1250;  
Best Local Similarity 21.2%; Pred. No. 6.4e-18;  
Matches 271; Conservative 187; Mismatches 452; Indels 368; Gaps 66;

QY 254 GRNGILGQWITSMTPAGLKETVYNNQGHFPQ-----SANLPVLPLYVTLMKQVPG 307  
DB 59 GMEGVAEQGLR-KARNAVEDAAEDASRQGAKEKDPVDVATGRWLPQI-----DVSIL 113  
QY 308 AQPAIQAEYSYTHNYVGGSGNGIWNKLDNLY-----GLMTEYNYGSTES 354  
DB 114 PGQLPLVWRQFESSYRLGWFPGTWSLTDQRLVDTAGVVLVGDGLVLAYPH---PA 170  
QY 355 RYKDKEGHQIVRIERTYNNVHLLTSECKQNGYIQTETAYAIIGHNFDSPQSQFOL 414  
DB 171 PGVPTLPSPGRWLSLDRITDGGYTL-----TDQKSGHIR----- 203  
QY 415 PKTKETWRSADNSYRSEITETTFDESNGPLTKVIXDKKTKQIISPSTHWEYYPAGEVD 474  
DB 204 -----HPVDRSDTLAVLEQLDDRNGWIT-----FEY-----DAD 233  
QY 475 NCP-----PEBYGTRFKV-----KIIQTPYDSEFKDDPEK-FIQYRYSLIGSQSHVTLKIEE 526  
DB 234 GAPRYLTHGGYRLRISTEAGRVTLHLASAADVGGDQLIRYGY-----TDGHL----- 283  
QY 527 RHYSATQLANST--LQYNTDKSELGRLLKQTE-----CTKGNGKTYSV 569  
DB 284 -----TEVNSGRPLQYVCD--ELGRITSWTDNDSHFSYAYDDEDRCTH-QSGAAGHL 335  
QY 570 VHKFTYTKOD-----DTLQOSHSLTTHDNFTIHRSQVRSRYTGLRFSDDTDKDIVTQM 622  
DB 336 RSTFAYGAVDPGTGAHTTVDYSYGQTH--YLINR-----RCQVIAETDALGAVTRY 386  
QY 623 SYDKLGRLLTRTLNSGTPYANTLYDY-ELNNLQDDNRPFPVITTTDVNGNQLRNEFDGA 681  
DB 387 QORDYNNLLSQT-----DPLGHTTSFRYDDAGNLVAATRP-----DGREARAENYAL 433  
QY 682 GRHVSQCLKSDGDKFYTHIQOYDEQGRHHTSTYSYDLNTRGROQTDPKVHLSMSKY 741  
DB 434 GLPVK--LVNPDG-----TITRQTFDERG-----NLTSVTDPS--GQTRFGY 472  
QY 742 DNWQJANTHWSGVSEKITVD-----PITLT-----ATKQLQSNNNVQTKSVTTTPSQ 793  
DB 473 DEGRUTSMTDPLGHTTGIVCORAGLPLVTDPLGAVTRYERDAFGRTTA--ITDPTGAT 530  
QY 794 QPTQITLPEAGHLQ-----SCHTLTRDGMWRVAKETDALGQCTIYQYDNNVRIQIT 846  
DB 531 TRLEWTV---EGHLSRTAPDGTSESWTYDGECCNCTSHTDPPGVSVLFEYTHFDLPTART 587  
QY 847 LPDGTIVNRKYAPFSDTLITDIRVNGISLGOQTFDGLSRLFQSQDQGRWYATYSAGND 906  
DB 588 GPDGV-----RY-EFEHD-----TELK-----LSQVTNPH--GLTNWYAYDAAGR 624

QY 907 QCPSTVITPDQGFIIHYQOPELDDAVLQVAS--NEITQOFSY-NPVTGALLKAVARGSQL 963  
DB 625 LVAETDF--DNRTLTYEY-----DFAGRLASRTNALGQMIAFERNELGQIVRKDAAGQAT 677  
QY 964 TPIIYPSGRLLKMNENDMKMSVLTWLRGLENGYTDLTGTGIQKISRDTGRVTVQIKDSSI 1023  
DB 678 TYAYDFTDQLAQ-----TQPDGTALTILDRHGRMISEAVNGR 716  
QY 1024 KTTLANYYDDLNRHI-----GSQVT-----DLA-----TGHMLTTTVTFDFGLNREIGRLCD 1068  
DB 717 ALTYDYDELGRTRRTTPSSATTINSYDVAGRTSMTSGRSIDFVYDEAGRELGRFF-- 774  
QY 1069 SSGHTLDIOOSWLKTOQLANRIVKLVQRTQEQYSYDSRNLNQYKCDGAECPDKYGH 1128  
DB 775 --GH-ITLHSPDALGRLTSSQV-LGPAGRRITQHRQY-----TYRAD-----GH 814  
QY 1129 SIVTQN-----FTYDIYGNITACHT-----TF-----ADGTEDHATFKFANP----- 1165  
DB 815 LIGIEDQLSGRRFDDLAGRVTVAVHVNVTETAYDAVGNQTSASWPAGHPQGAIGNR 874  
QY 1166 ----TDPCLQTEVHHTHPDM-----PDNIRLKVDKAGRVINITDNHGVTEN 1207  
DB 875 TTYGSLTRAGQVRVEHDLGRLILLRQKTRLSRKPDTWYTWDAEDRLTSVTPDGTWR 934  
QY 1208 FTYDTLGR-----LQNGQSGVYGYDPLNRLVSKQTDLDCELYYRETMVNEVR--N 1257  
DB 935 YTVDFLGRRTAKRLAEDGETLV-----ERVDFTWDGTVLCBQTTTSPDLPHQVTLTWD 988  
QY 1258 GEMIRLLRTGETIIAQOARSKVL-----LTGTDQQSVILTSKQNLQSOEAYSAY 1307  
DB 989 HOGRLPVTQTERIVAADAPQOQIDSRFFAIVTDLVGTPSE---LLDERGEIVWRTATLW 1045  
QY 1308 GKHKSTANDASI--LYNGERADPVSGVTHLNGVRSYDPTLMRFTPDSLSPPGAG-GI 1364  
DB 1046 GSTTWAKNSTATPLRFPQYYPETGLHY--NYFRHYDETARYVTPD---PLGLGPAP 1100  
QY 1365 NPYCYCLGDPINRSPSG 1382  
DB 1101 NPAAY-VPNPHMWADPLG 1117

## RESULT 6

US-10-156-761-7751  
; Sequence 7751, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7751  
; LENGTH: 2386  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-7751

Query Match 4.4%; Score 391.5; DB 14; Length 2386;  
Best Local Similarity 20.1%; Pred. No. 1.9e-17;  
Matches 379; Conservative 214; Mismatches 688; Indels 601; Gaps 93;



;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
;; FILE REFERENCE: ELITRA.034A  
;; CURRENT APPLICATION NUMBER: US/10/282,122A  
;; CURRENT FILING DATE: 2003-02-20  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/230,335  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/230,347  
;; PRIOR FILING DATE: 2000-09-09  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/267,636  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 78614  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 46565  
;; LENGTH: 2234  
;; TYPE: PRF  
;; ORGANISM: Bacillus anthracis  
;; US-10-282-122A-46565

Query Match 4.3%; Score 386; DB 15; Length 2234;  
Best Local Similarity 18.7%; Pred. No. 4.2e-17;  
Matches 339; Conservative 236; Mismatches 633; Indels 608; Gaps 76;  
  
QY 21 FTQANNFTSAVSGVDPR-----TGLYNIQTTLGHIVG-----NGNL----- 57  
DB 748 YSKVGHAESEISHEFVPRFSQDSGFGTFMFDYASIPVLGNKVATNGNFMSEKDIITLSG 807  
  
QY 58 -GPTLPLTSLYSPLNKTDIGFGFNPLSV-----YDRKNSLLSL-----TGEN- 102  
DB 808 RGPDSVSVERTYNSQSKVGLFGTWSGLEERVMADGNGNLLIISTDGANITFTRTGDNK 867  
  
QY 103 -----YKVIETDKTVKLQOKKDLNLRPEKDLKENCYRIIHKSGDIEVL 145  
DB 868 YQAPTGIYILEIKVSGGYEIKDKQIV-----TFYKSGDAQGR 905  
  
QY 146 TGFNNNAFDLQVP-----KKLLNPAGHAIYIDWNFEATQPRLNRIYDLDLGHDIPL 196  
DB 906 IEVTKDKYGNNTTYEYDYGASRLSKVKNASGKELVL--QYDGNKKAAARVI---GPDNKT 959  
  
QY 197 LNLVYQGLKITLTLPPGQ--KEGYTELRFLNRLQNSIHNSLGNENPLTWSFGY---- 250  
DB 960 ITFNVDGDLVLSSTTPEGKVIKGYD-----NGVLTSIYDPOHTDAKPKYTSAYENDR 1013  
  
QY 251 -----TPICKNGIL-----GOWITSMTPAGGLKETVYNSNNQGHFPQSANLPLVPYVTL 301  
DB 1014 LVKVTDLPGKATTLAVNTSGKEVTLNPKRKVTYIND-----AGNPV----- 1057  
  
QY 302 MKQVPGAGQPAIQAEYSYTSYSHNVYVG-----GSGNIGWNNKLDNLYGLMTEYN- 348  
DB 1058 -KTVEDVGRNLNTTSYEYNNANLVKTTTPKNQOTETATYDNGNVTSTVDEMGTEKFEYNK 1116  
  
QY 349 -YGSTESRRYKKE-----GHDQIVRIERTVNNHLLITSECKQONGVIQTETAYAI 400  
DB 1117 DNGIILKATNEDRKTTVAVVGANTEVSQTDQGAN-----TSSVIHHDQYGNPIETSKELS 1171  
  
QY 401 IGHNFDSQSQPLPKTKTETWRSADNSYRSEITETTFDESNGNPLTKVKDKKTKIISP 460  
DB 1172 AGGNLQNPS-FEM--NGTERKWKVDTNNSGSIS-----KDATPAPGGLGESSLKITTK 1223

QY 461 STH--WEYPPAGEVDCPPBPYGFTRFKKII-----QTPYDSEPKDDP 503  
DB 1224 ATNDMGYIAAIOEVTLEPNTTYTLSCMVKTDLVNGAAFFNVQSLNENGAGIDGWHDR 1283  
  
QY 504 EKFIQYRSLIGSQSHVTLKIBERHYSATQLLNSTLFOYNTDKSELGLL-LKQTECTKGE 562  
DB 1284 HNKVQGTSDWNRQ--VTFKTEQ-----TRKVIYLOVENGGSATSGSAWFDKIQLEKGE 1337  
  
QY 563 NGKTSYVHKFTYTKO--DDTLQQ-SHSITTH-----DNFTIHRQSVRSRYTGLRFS 611  
DB 1338 VSSSFNPVLNSSFENWPDGFPQWVRSCSQHCERNDVSDSFTGHSSIVMER-----S 1391  
  
QY 612 DTDTKDIVTOMSYDKLGRLLTRTLNSGTPYANTLTVDYELNLLNODNRPPFVITTTDVG 671  
DB 1392 EYGPNDI-----GYRNRVILNQKAEVTLTAMSKSENVDN----- 1428  
  
QY 672 NQLRNEFDGAGRHSVQCLXDSGDGKFYTHIQOYDEQGRHHTSTSYDYLNTNGRQOTDP- 730  
DB 1429 -----APDKLSKDYAVLAETYYQDG-----TVVNYTSPSGTNDW 1464  
  
QY 731 -----DKVHLSMSKSYDNWQIANTHWSYGVSEKITVDPITLTATKQLQSN 777  
DB 1465 NRSAAVIPAAPKPIQKIEIFLLFRKNNKG-----KWFFDIRLEGNALIKNE 1511  
  
QY 778 NNVTQKEVTVTPSQOPIQITLFDAGHLQSCHTLTRDQGWDRVKRETDAIGQCTIYQYD 837  
DB 1512 YD-NDGNVVATY-----DREGQK--NTFTYDASGNKKSETDEKGNTKLYDYN 1555  
  
QY 838 NYNRVIQITLPDGTIVNRKYAPSTDTLTITDTRVNGISLGQOQTFDGLSRLTQSDGGRVW 897  
DB 1556 KDNLLTKVTLKNGTSVNYRYD----- 1576  
  
QY 898 AVTYSAGNDQCPSTVITPDGQFIHYOYQPELDDAVLQVASNEITQOFSYNPVTGALLKAV 957  
DB 1577 -----HNGNTEKSVMPGGKTQTHKYEYDVNKNVTVIDALNRRIEN-TYDENANKIKTKM 1631  
  
QY 958 ASGQSLTPIYYPGRLKMNINDMKMSYLTWLRGLENGYDITGTIQIKISRTHGRVTQ 1017  
DB 1632 PNGSILESVDYADRVRVGEK-----RNGKDSFT-----PERDQNGQVTK 1670  
  
QY 1018 IKD--SSIKTLNVDLNRHIGSOVTDLATGHMLTTTVFDFGLNRIGRKLCDSSGHTLD 1075  
DB 1671 VKDLVNGVERTKTKADAR-----VTS-AT-----DSRGKID 1702  
  
QY 1076 IQQSWL-----KTQOLANRIVKLGVLQRTQEQSYSDSRNRLNQ--YKCDGAECPCTDKY 1126  
DB 1703 -----WAYHDKANSKTEKLEQIVTQGGY---TNKVSYD-YNTLDQNRIVTDGSG- 1748  
  
QY 1127 GHSIVTQNFYDIYGNITACHTTTFADGTEDHATFKF--ANP-----TDPQOLTEVHHT 1177  
DB 1749 -----TYRFDYDDQGNV-----RTYTAGSGSGSTFNVDQANKIKDLVVGTSNILLSERYE 1799  
  
QY 1178 HPDMPDNILKYDKAG-----RVINITDNHNTENFTYDILGR-----LQ 1217  
DB 1800 YDQSGNRTKIKHEGAGGKVTETNFVYDPINQLLNEVLNPGTKTSYTDGFGNRTSVKIVIE 1859  
  
QY 1218 NQOGS---VYGYDPLNRLVSQKTDITLDCEL----- 1244  
DB 1860 NGKETSIAATFNEGQLVKFGNESLTYDVNGNRSTGKYKYTNWEDDQIVAITKQGENN 1919  
  
QY 1245 -----YYRETLVNEVRNEMIRLLRTGETI-----IAOQ 1274  
DB 1920 AFATKYDEDNRRIEKNVNGQVTRFYDGDGSIINPLVETDNGTVLRQYVVSADGARLAMK 1979  
  
QY 1275 RASKVLLTGTDSQQSVI---LTSQKQNLQSOEAYSAGK-----HKSTANDASILGNGERA 1327  
DB 1980 AOGQTLXYHNPGRDGVAMTNQDKEVATYEYDAWGNVLTSDTKGIAAD-NPEFYAGYMY 2038  
  
QY 1328 DPVSGVTHLNGVRSYDPTLMEFHTPD-----SLSPFAG-----GINPYSYCLGPPINR 1377  
DB 2039 DKEIGMYIL-----IARYNPEHGVFLVSDPDGDEDDPVTMNGYVTDADNPNVMM 2088  
  
QY 1378 SDPSGHLNQWATGIGWGIAGLLLTATGMAIAAGGIAAAIASTSTTALAFGALSPTS 1437







[illegible]

```
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43059
; LENGTH: 1397
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43059

Query Match      3.9%; Score 342; DB 15; Length 1397;
Best Local Similarity 21.4%; Pred. No. 2.4e-14;
Matches 256; Conservative 157; Mismatches 382; Indels 404; Gaps 63;

QY 418 KTEW-RSADNSYRSEITETTESGNPLTKVI-----KDKTKTKIIS-----PS-----461
DB 218 RTQTFHREAAGEFSGELTGVTDAGRHFLVLTQQAERARQQAISGGTSPAPDPT 276
QY 462 -----THWEYPPAGEVNDCCPEP-----YGFY-----RFLVK 488
DB 277 LPGVTEYGRDNGIRLSAOWLTHDPEY-----ENLPAAPLVRYGWTFRGELAANVYDRSNT 331
QY 489 KIIQTPYDSEFK-----DPEKFIQYRSLIGSQSHVTLKIBERHYSATQOLLNSTL 539
DB 332 QVASFTYDDKYGRMVAHRHTGRPE--ICRYD---SDGRVTQLNPAGLSYT-----379
QY 540 FOYNTKSELGRLLKQTEC--TKGNGKTYSVVHKFTYTKQDDTLOQSHSIITHDNFTIH 597
DB 380 YOYEKDRITITDSLRNREVLHTQEGG-LKRVVKK-----EHADGSVT 421
QY 598 RSQVRGRYTRGLPDSFTDKDIVTQMSYDKLGRLLTRTLNSGTPTVANTLTVDVELNLQDD 657
DB 422 QSQFDA--VGLRAQTDAAAGRTTEYSFDPVVTGLITRIT---TPDGRASAFY-----468
QY 658 NRPPFVITTDVNGNOLRNEFDGAGRHSQCLKSD-----GDGKFYTIHTQYDQGRH 712
DB 469 NHHSQLTSATGPDGLREYDEWGLRIOETAPDGDITRYDNPDSLPCATEDATGSR 528
QY 713 HSTYSYDLTNGHQDTPDKVHLSMSKSYDNGQIANTHWSYGVSKITVDPT-TLTATK 771
DB 529 KMTWISRY---GQLLSFTDCSGYVTRYDHRFQQTAVHREGLSQYRAYDSRGQLIAYK 585
QY 772 QLQSNMNVQTKGEVITYTPSQOPIQITLFDAGHLSQCHTLTRDCMDRVKRTDAIGO- 830
DB 586 DTQGHETRYE-----YNAAGDLTT--VIAPDG-SRNGTQYDANGKA 623
QY 831 -CTI-----YOYDNNRVQITLPGTIVNRKYAPFSTDTLITIRVNGISLGOQTF- 881
DB 624 ICTTQGLTRSMEDYDAAGRIVLTSENGS-----HTFRYDVLRLIQTGPDGRTQRYH 678
QY 882 -DGLSRLTQSDQGRWAWTYTSAGNDQCPSTVTPDGFHYQYQBELDDAVLQVASNEI 940
DB 679 HDLTGKLIRSEDEGLV-----THWYD-EADRLTHRTVNGET 714
QY 941 TQOFSNPVTGAL--LKAVAEGSLTPY-YPS-GRLKMENI-----NDM---KKMSY 986
DB 715 AERWQYDE-RGWLTDISHSEGRHVTVHYGYSKGRLESHLTVHPQTNELLMQHETRH 773
QY 987 LMTLRGLENGYDLDITQIKISRDTHGRVTQIKDSSIKITLNYDLNRHIGSOVTDLATG 1046
DB 774 AYNAQGLAN-RCIPDPLPAVEMWTY-----SGWLSOMKLG 808
QY 1047 HMLTTTVER--DGLNREIGRKLCDSSGHTLDIQOSWLKTOQLANRIVKLVNGVLRTEQYS 1104
DB 809 D--TPLVEYTRDLRHRETLR-----SFGRYELTAVTPAGLQSQ--HLNSLLS-DRDYT 858
QY 1105 YDSRNLNQYKDCGACPTDKYHSIVTQNTFYDIYGNITACTHTPAD-----GTEDHA 1158
DB 859 WNDNGELIR-----ISSPRQ-----TRSYSYSTTGRLTGVHTTAANLDIRPYTTPDA 906
QY 1159 TPKFANPTDPCQLTEVHHHTHPD-----MPDN-----IRLKDYKAGRVINITD-----1200
DB 907 GNRLLPDP-----ELHPSALSWMFNDRIARDAHYLYRYDRHGRTEKTDLIPIGEV 956
QY 1201 ---NHGNTENFTYDTLGRQLONGQGSVYG-----YDPLNRLVSQKTDTLDCEL-----1244
```

```
Query Match      3.8%; Score 340.5; DB 15; Length 1411;
Best Local Similarity 20.4%; Pred. No. 3.1e-14;
Matches 287; Conservative 172; Mismatches 459; Indels 490; Gaps 70;
```



Db 475 --GNQLTAVVSPD--GLESREYDEPRLV-----SE 502

Qy 779 NVOTGKEVT--TYTPSQOPIQITLDFEAGHLQSCHTLTRDGDWRKVKETDAIGQCTIYQYD 837

Db 503 TSSRGVTRYRYDAHSELPAITTDATG--STRQMTWSRYGQLLAFTDCSGYOTRYEYD 559

Qy 838 NYNRVQITLTPDGTIVNRKYPASTDLTDIRVNGISLQOQTFDGLSLRLTQSDG-GRV 896

Db 560 RFGOM-----TAVHRE-----EGISL--YRRYDNRGLTSLVKDAQGRE 595

Qy 897 WAYTSAGNDQCSTVITPDGQFIHYQY-----925

Db 596 TRYENAGDL--TAVITPDGNSETQYDANGKAVSTTQGLTRSMEDYDAAGRVISLTNE 553

Qy 926 -----PELDDAVLQVANSBITQFQSNYPVTGALLKAVAQOSLTPPIYPSGRLKME 976

Db 654 NGSHSVFSDALDRLVQGGFDQRTQRYHYD-LTGKLTQSEDEGLVILWYDESDRITHR 712

Qy 977 NINDMKMSYLWLRGLENGYTDLTGTIQKISDTHGRVTOIKDSS-----IKTILNYDDL 1032

Db 713 TVNGEPAEQWY-----DGHGWLTDISHLSEGRVAVHYGYDDK 751

Qy 1033 NRHIG--SOVTDLATGHML-----1049

Db 752 GLTGEQOTVENPETGELLWQHETKAYNEOGLANRVTPDSLPPVWLTYSGLYAGMKL 811

Qy 1050 --TTTVEP--DGLNREIRKLCDSGH--TLDIQOSWLKTQQLANR-----1089

Db 812 GGTPLVEYTRDLRHRETVRSFGSMAGSNAAYELSTYTPAGQLQOSHLNSLVYDRDYGS 871

Qy 1090 ----IVKLVGLORTQSYSDSNRLNQYKCDGAC-----PTDKYHGI-----1130

Db 872 DNGDLVRISGPRQ--TREYGYSATGRLESVRTLAPDLDIRPYATDPAGNRLDPPELHPDS 930

Qy 1131 -----VTQN-----FYDIYGNLTACHTTFADG--TEDHATKFPANPTDPCOLTEV 1174

Db 931 TLTWPDNRDAEDAHVYRHDEYGRTEKTDRIAPAGVIRTDERT-----975

Qy 1175 HHTHPDMPNIRLKYDKAGRVINITD--NHGN---TENFYDTIGR-----1215

Db 976 HHVH-----YDSQHLVFTYTRIQHGEPLVESRYLYDPLGRMAKVRWRERDLTG 1025

Qy 1216 ----LQNGQSVYGYDPLNRLVSQKTDLDCELYR-----1247

Db 1026 WMSLSRKPEVTWYGDG-DRLATVQDTRIQTVYEPGSGFTPLIRVETENGEREKAQRS 1084

Qy 1248 --ETMLVNRNG-----EMIRLL-RTGETIIA-----OQBASKV 1279

Db 1085 LAETLQOEGSENGHVVFPAELVRLDLREEIRADRVSESGRAWLAQCGLTVEQLARQV 1144

Qy 1280 LLTGTDSQOS-----VILTSDKONLSQEA--YSAVGKHKSTANDASILGYN---1323

Db 1145 EPEYTPARKAHLYCHDRGLPLALISEDGNTAWSAEYDEWGNQLNENPHV--YQPYRL 1202

Qy 1324 -GERADPVSGVTHLNGYRSYDPTLMRFPHTPDSLPFG--AGGINPYSYCLGPPINRSDPS 1381

Db 1203 PGQOHDDEBSGLY--NRHRYDPLQGRYITQD---PMGLKGGWNLXYQYEL-NPLQIDPM 1256

Qy 1382 GHLISQAWTIGMGINGIAGLLLTATGMAIAAGGIAAAATSTTALAFALSVTSDITS 1441

Db 1257 GLL--QTWDDARSG-----ACTGGV---CGVLSRIIGPSKFDSTADAALD-----1296

Qy 1442 IVSGALEDASPKASSILGVSMGMGAAGLAESAIGKGTKLATHLGAFAEDGNALLKSTS 1501

Db 1297 ----ALKETQNRS-----LCNDMEYSGIVCKDTNG-----KYFASKAETDNLK--K 1336

Qy 1502 ESSRIKWGTVRSLDREIVRNEESQVVKHRSRGYTDNFMGKGBOAILVHGDK--DGFLYHT 1559

Db 1337 ESYPLKRCPTGTRVAAVYHTG---ADSHGDYVDEFFSSSKNLVRSKDNLEAFYLAT 1393

Qy 1560 EGNKH---NGKGPY 1570

Db 1394 PDGRFEALNNKGEY 1407

## RESULT 14

US-10-282-122A-42617  
; Sequence 42617, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 42617  
; LENGTH: 1426  
; TYPE: PR1  
; ORGANISM: Escherichia coli  
US-10-282-122A-42617

Query Match 3.8%; Score 339; DB 15; Length 1426;  
Best Local Similarity 20.2%; Pred. No. 4.1e-14;  
Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;

Qy 572 KFTYTKQDDTLQSSSIHTHNFTHRSQVRSRYTGRFSFDDTDKIVTQMSYDKLGRLL 631

Db 317 RYITTEAGELL-----AVDRSNTQVRAFYTDAQHPGFMVAHRYAGRPEMYRYDDTGRVV 372

Qy 632 TRLNSGTPYANTLTYYELANNLQDDNRPFFVITTTD-----668

Db 373 EQLNPAG-----LSYRY-----LYEQDR-----ITVTDLSNRREVLHTEGAGLKRUVKKE 418

Qy 669 -VNGNQLNREFPAGRHVSQCLKDSGDGKFTYHTHQYD-----EQGRHHTSYSDYL 721

Db 419 LADGSVTRSGYDAAGRLTAQ--TDAAGRRTEYGLNVWSGDIITDITPDGRETGFYND-- 474

Qy 722 TNGRQOT---DPDKVHLSMSKSYDNWQIANTHWSYGVSEKIVDPITLTATKQLQSN 778

Db 475 --GNQLTAVVSPD--GLESREYDEPRLV-----SE 502

Qy 779 NVOTGKEVT--TYTPSQOPIQITLDFEAGHLQSCHTLTRDGDWRKVKETDAIGQCTIYQYD 837





Db 414 -----EHADGVSQSFDA--VGRLAQTDAAGRTEYSPDVVTGLITRIT---TPDGR 462  
QY 644 TLYDYELANLQDNRPPFVITTTDVNGQLRNEFDGAGRHSQCCLKSD-----GDGRF 698  
Db 463 ASAFYNNHNQ-----LTSATGPDGLREYDELGRLIQETAPDGDITRYRYDNPH 514  
QY 699 YTIHTQOYDEQGRHHTSTYSYDLTNGRQQTDPKVKHLSMSKSYDNWGOIANTHWSYGVSE 758  
Db 515 SDLPFCATEDATGSRKMTWSRY---QQLLSFTDCSGYVTRYDHDREFGQMTAVHREELSQ 571  
QY 759 KITVDPI-TLTATKQLQSNNNVQTGKEVTTVTPSQPIQITLFDAGHLQ-----SC 810  
Db 572 YRAYDSRGQLIAVKDTQGHETRYE-----YNIAGDLTAVIAPDGR 612  
QY 811 HTLTRDGDWRVRKETDAIQCTIYQDYNRVIQITLPDGTIVNRKYAPFSTDTLTIDIR 870  
Db 613 NGTOYDAMGKAVRTTQG-GLTRSMEDYDAAGRIVRLTSENGS-----HTTFRYDVLDRLIQ 666  
QY 871 VNGISLGOQTF--DGLSRLTQSDQGRVWAYTYSAGNDQCPSTVITPDGQFIHQYQPEL 928  
Db 667 ETGFDGRTQRYHDLTGKLI RSEDEGLV-----THWHYD-EA 702  
QY 929 DDAVLQVASNEITQOFSYNPVTGAL--LKAVAGQSILTPIY---PSGRL--KMNINDM 981  
Db 703 DRLTHRTVGETAERWQYDE-RGWLTDISHISEGRVA-VHYRYDEKGRLTGERQTVVHP 760  
QY 982 KMSYILW-----TLRGLNGYDITGTIQKISRDTGHRVTQIKDSSIKITTLNYYDDL 1033  
Db 761 QTEALLWQHETHRAYNAQGLAN--RCIPDSLPAVWELTYG-----SCYLAGMKLGD-- 809  
QY 1034 RHIGSQVTLATGHMLTTTVEP--DGLNREIRKGLCDSSGHTLDDIOQSWLKTQQLANRIV 1091  
Db 810 -----TPLVEYTRDLRHRETLR-----SFGRYELTAYTPAGOLQSQ-- 846  
QY 1092 KLVGLQRTQOYSYDSRNLQYKCDGASCPTDKYCHSIVTQNFYDIYGNITACHTTEA 1151  
Db 847 HLNSLLS--DRDYTNWNGELIR-----ISSPRQ-----TRSYSTTGRLTGVHTTAA 893  
QY 1152 DGTEDHATEKFANPTDPC--QLTEVHHHPD-----MPDN-----IRLKYDKAGRVIN 1197  
Db 894 -----NLDIRIPYATDPAGNRLPD-PELHPDSTLSMWPDRNRIARDAHYLYRYDRHGRLTE 947  
QY 1198 ITD-----NHGNTENFTYDTLGRLONGQ-----GSVGYDPLNRLVSKTDT 1239  
Db 948 KTDLIEGVIRTDDETRHYDQSHRLVHYTRTOYEELVESRYLYDPLGRRVAKRVMR 1007  
QY 1240 LDCEL-----YY-----RETMVNE-----VNGEMIRLLR-----TGETI 1270  
Db 1008 RERDLTGWMSLRKPKQVWYGMGDRLTIONDRTRIQTIOYQGSFTPLIRVETATGELA 1067  
QY 1271 IAQORA-----SKVLLTGTDSQOSVIL-----TS 1294  
Db 1068 KTORSLADALQOSGEDGSGVVFPVLVQMLDRLESEILADRVSEESRRWLASCGLTVE 1127  
QY 1295 DKQN-----LSQEA-----YSAYGKHKSTAND---AS 1318  
Db 1128 QMOMQMDPVYTPARKIHLVCHDRGLPLALISKEGTTWCABEYDEWGNLNEENPHOLQQ 1187  
QY 1319 ILGYNGERADPVSGVTHLNGVRSYDPTLMRFHTPDSLSPPG-AGGINPYSYCLGDPINR 1377  
Db 1188 LIRLPQQQYDEESGLY--NRHRYDPLOGRYITQD---PIGLKGNFYQYPL-NPVTN 1241  
QY 1378 SDPSG 1382  
Db 1242 TDPLG 1246

Search completed: February 9, 2005, 18:23:27  
Job time : 197 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2005, 17:59:19 ; Search time 206 Seconds  
(without alignments)  
3141.020 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879

Sequence: 1 VVIKFLKLFRRITMSDNEF.....PRKIILGRTEKTVKPTFRP 1673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: A\_Geneseq\_16Dec04:\*
- 2: Geneseqp1980s:\*
- 3: Geneseqp1990s:\*
- 4: Geneseqp2000s:\*
- 5: Geneseqp2001s:\*
- 6: Geneseqp2002s:\*
- 7: Geneseqp2003as:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8879	100.0	1673	3	Aay95707 Cosmid cH
2	1638	18.4	1584	2	Aay33727 Photorhab
3	1601	18.0	1410	8	Adg26337 Chromobac
4	1583	17.8	1787	6	Abm67433 Photorhab
5	1569	17.7	1590	6	Abm67283 Photorhab
6	932.5	10.5	982	7	Adg73113 Pseudomon
7	932.5	10.5	982	7	Adl12168 Pseudomon
8	517.5	5.8	1317	8	AdS22210 Bacterial
9	432	4.9	2334	5	Abg31849 Human kin
10	386	4.3	2234	6	Abu18641 Protein e
11	365	4.1	1400	7	Adc01365 Enterorhae
12	361	4.0	1510	7	Adg30698 Xanthomon
13	353.5	4.0	843	6	Abu48290 Protein e
14	350	3.9	1515	6	Abu19676 Protein e
15	342	3.9	1397	6	Abu15135 Protein e
16	341.5	3.8	1404	7	Adc00960 Enterorhae
17	341	3.8	1394	7	Adc01426 Enterorhae
18	340.5	3.8	1411	6	Abu15136 Protein e
19	339	3.8	1426	3	Abu15983 E. coli p
20	339	3.8	1426	6	Abu14693 Protein e
21	331.5	3.7	1377	4	Aau34791 E. coli c
22	331.5	3.7	1377	4	Aag98997 E. coli c
23	331.5	3.7	1377	6	Abu14807 Protein e
24	328	3.7	1572	6	Abu41491 Protein e
25	324.5	3.7	1429	6	Abm69806 Photorhab

#### ALIGNMENTS

RESULT 1

AAY95707

ID AAY95707 standard; protein; 1673 AA.

XX AC AAY95707;

XX DT 25-OCT-2000 (first entry)

XX DE Cosmid CHIRMS encoded protein P14-2f.

XX KW Cosmid CHIRMS; nematocide; nematode; biological control agent; transgenic plant; helminthiasis; P14-2f.

XX OS Xenorhabdus bovienii.

XX PN WO200042855-A1.

XX PD 27-JUL-2000.

XX PF 24-JAN-2000; 2000WO-CB0000219.

XX PR 22-JAN-1999; 99GB-00001499.

XX (HORT-) HORTICULTURE RES INT.

XX Morgan JAW, Jarrett P, Ellis D, Ousley MA;

XX WPI: 2000-499157/44.

XX N-PSDB; AAA50029.

XX Novel composition used to control parasitic nematodes, especially in a plants such as maize, cotton, soya, and rice, comprises a bacterium which is a symbiont of an entomopathogenic nematode.

XX Example 6; Page 42-43; 74pp; English.

XX The present sequence is that of protein P14-2f encoded by an open reading frame identified in cosmid CHIRMS (see AAA50029). CHIRMS was obtained by ligating Xenorhabdus bovienii strain I73 (NCIMB 40986) Sau3A-digested DNA fragments into the BamHI site of the Strategene cosmid vector Supercoils, packaging into Escherichia coli XL Blue 1, and screening for nematocidal activity against Caenorhabditis elegans. Analysis of the DNA indicated a number of open reading frames for which the corresponding protein CC sequences were determined (see AAY95685-Y95735). Nematodes can be CC controlled through the use of bacteria associated symbiotically with an CC entomopathogenic nematode. Such bacteria include Xenorhabdus and CC Photorhabdus spp. such as X. bovienii strain I73. The symbiont bacteria, CC an engineered bacterium, or a nematocidal protein obtained from such

26	322	3.6	1385	6	ABU40318	Abu40318 Protein e
27	317.5	3.6	1504	6	ABM66973	Abm66973 Photorhab
28	317.5	3.6	1565	6	ABU16634	Abu16634 Protein e
29	317	3.6	998	7	ABO79678	Abu79678 Pseudomon
30	314	3.5	1395	6	ABU22662	Abu22662 Protein e
31	313	3.5	1627	6	ADA35317	Ada35317 Acinetoba
32	307.5	3.5	932	6	ABU16693	Abu16693 Protein e
33	306.5	3.5	1530	6	ABU40251	Abu40251 Protein e
34	306	3.4	1586	7	ADF05044	Adf05044 Bacterial
35	298.5	3.4	1531	6	ABU21384	Abu21384 Protein e
36	298	3.4	885	4	AU52772	Au52772 Propionib
37	298	3.4	885	6	ABM49291	Abm49291 Propionib
38	298	3.4	920	6	ABM65127	Abm65127 Propionib
39	297	3.3	1512	6	ABU50462	Abu50462 Protein e
40	294	3.3	1439	7	ADF07275	Adf07275 Bacterial
41	292	3.3	1626	7	ABO75059	Abu75059 Pseudomon
42	291.5	3.3	1253	6	ABU40990	Abu40990 Protein e
43	291	3.3	2346	5	ADI16955	Adi16955 Marine NO
44	290.5	3.3	2802	5	ADI16953	Adi16953 Chicken N
45	289.5	3.3	1364	6	ABU47385	Abu47385 Protein e

CC bacteria, particularly p13-1f (see AAY95706) or P14-2f can be used to  
CC control helminthiasis in a human or domesticated animal or for the  
CC control of plant pathogen nematodes. Also claimed are vectors for  
CC expressing nematocidal proteins in host cells, and transgenic plants  
XX  
SQ Sequence 1673 AA;

Query Match 100.0%; Score 8879; DB 3; Length 1673;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIEKLKLRITMSDNNNEFFTQANNFTSAVGGVDPRGLYNIQITLGHVGNLQPT 60  
DB 1 VVIEKLKLRITMSDNNNEFFTQANNFTSAVGGVDPRGLYNIQITLGHVGNLQPT 60

QY 61 LPLTISYSPLNKTDIGFIGNFGLSVYDRKNSLLSLSTGENYKVIETDKTKLOQKKLD 120  
DB 61 LPLTISYSPLNKTDIGFIGNFGLSVYDRKNSLLSLSTGENYKVIETDKTKLOQKKLD 120

QY 121 NLRFEKDKENCYRIIHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDWNPETAP 180  
DB 121 NLRFEKDKENCYRIIHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDWNPETAP 180

QY 181 RLNRYYDDLDGHDIPLLNLEYQGLIKTILTFPGQKEGYRTELRFNLRLNSIHNFSLGN 240  
DB 181 RLNRYYDDLDGHDIPLLNLEYQGLIKTILTFPGQKEGYRTELRFNLRLNSIHNFSLGN 240

QY 241 ENPLTWSFGYTPIGKNGILQWITSMAPGGLKEFTVYNNQGHFFQSANLPVLPYVT 300  
DB 241 ENPLTWSFGYTPIGKNGILQWITSMAPGGLKEFTVYNNQGHFFQSANLPVLPYVT 300

QY 301 LMKQVPGAGOPAIQAEYSVTSNNYVGGSGNGIWNKLDNLYGLMTEYNGSTESRYKDK 360  
DB 301 LMKQVPGAGOPAIQAEYSVTSNNYVGGSGNGIWNKLDNLYGLMTEYNGSTESRYKDK 360

QY 361 EGHQDQIVRIERTYNNYHLLTSECKQNGYIOTTETAYYAIIGHNFDSPSQPLPKTKTE 420  
DB 361 EGHQDQIVRIERTYNNYHLLTSECKQNGYIOTTETAYYAIIGHNFDSPSQPLPKTKTE 420

QY 421 TWRSDNSYRSITETTFDESNGPLTKVKKTKTKIISPSHWEYYPAGEVDCNCPPEP 480  
DB 421 TWRSDNSYRSITETTFDESNGPLTKVKKTKTKIISPSHWEYYPAGEVDCNCPPEP 480

QY 481 YGFTREVKKIIOTPDSEBKDPKEFIOYRSLIGSQSHVTLKIBERHYSATQLLNSTLF 540  
DB 481 YGFTREVKKIIOTPDSEBKDPKEFIOYRSLIGSQSHVTLKIBERHYSATQLLNSTLF 540

QY 541 QYNTDKSELGRLLKQTECKGNGKTYSVVHKFTYTKQDDTLQOQSHSITTHDNFTIHRSQ 600  
DB 541 QYNTDKSELGRLLKQTECKGNGKTYSVVHKFTYTKQDDTLQOQSHSITTHDNFTIHRSQ 600

QY 601 VRSRYTGRFLPSDTHDKDIQVMSYDKLGRLLTRTLNSGTPYANTLTIDYELANLQDDNRP 660  
DB 601 VRSRYTGRFLPSDTHDKDIQVMSYDKLGRLLTRTLNSGTPYANTLTIDYELANLQDDNRP 660

QY 661 PFVITTTDVGQNLRFEPDAGRHVSQCLKDSGDGKFYTIHTQQYDEQGRHHTSYSDY 720  
DB 661 PFVITTTDVGQNLRFEPDAGRHVSQCLKDSGDGKFYTIHTQQYDEQGRHHTSYSDY 720

QY 721 LTNRGQRTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKIVDPIPLTATKQLQSNVNV 780  
DB 721 LTNRGQRTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKIVDPIPLTATKQLQSNVNV 780

QY 781 QTGKEVTTTPSOQPIQITLFEAGHLQSCHTLTRDGDWRVRKETDAIGCQTIYQYDYNV 840  
DB 781 QTGKEVTTTPSOQPIQITLFEAGHLQSCHTLTRDGDWRVRKETDAIGCQTIYQYDYNV 840

QY 841 RVQITLPDGTIVNRKYAFSTDTLITDIRVNGISILGQOTFDGLSRLTQSDGGRWYAT 900  
DB 841 RVQITLPDGTIVNRKYAFSTDTLITDIRVNGISILGQOTFDGLSRLTQSDGGRWYAT 900

QY 901 YSAGNDCQSPSTVTPDQGFHYQVPELDDAVLQVASNEITQOFSNPNVTGALLKVAEG 960  
DB 901 YSAGNDCQSPSTVTPDQGFHYQVPELDDAVLQVASNEITQOFSNPNVTGALLKVAEG 960

DB 901 YSAGNDCQSPSTVTPDQGFHYQVPELDDAVLQVASNEITQOFSNPNVTGALLKVAEG 960  
QY 961 QSLTPIYPSGRKWMENINDMKWMSYLWTLRGLNGYTDLTGTIQKISRTHGRVTOIKD 1020  
DB 961 QSLTPIYPSGRKWMENINDMKWMSYLWTLRGLNGYTDLTGTIQKISRTHGRVTOIKD 1020

QY 1021 SSIKTTLYDNLNRHIGSQVTDLATGHMLTTTVEFDGLNREIRGRKLCSSGHTLIDIQOSW 1080  
DB 1021 SSIKTTLYDNLNRHIGSQVTDLATGHMLTTTVEFDGLNREIRGRKLCSSGHTLIDIQOSW 1080

QY 1081 LKTOOLANRIVKLVNGVLRQTEQSYSDSRNRLNQKCDGAECPDTKYGHISIVTONFTYDIY 1140  
DB 1081 LKTOOLANRIVKLVNGVLRQTEQSYSDSRNRLNQKCDGAECPDTKYGHISIVTONFTYDIY 1140

QY 1141 GNIITACHTTADGTEDHATFKFANPTDPCOLTEVHHTHPDPMPDNIRUKYKAGRVINITD 1200  
DB 1141 GNIITACHTTADGTEDHATFKFANPTDPCOLTEVHHTHPDPMPDNIRUKYKAGRVINITD 1200

QY 1201 NHGNTENFTYDTLGRLQNGQSVYGYDPLNRLVSQKTDTLDCELYRETMLVNEVRNGEM 1260  
DB 1201 NHGNTENFTYDTLGRLQNGQSVYGYDPLNRLVSQKTDTLDCELYRETMLVNEVRNGEM 1260

QY 1261 IRLRTGETIIAQORASKVLLTGTDSQOSVILTSQKNSQEAAYSAYGKHKSTANDASIL 1320  
DB 1261 IRLRTGETIIAQORASKVLLTGTDSQOSVILTSQKNSQEAAYSAYGKHKSTANDASIL 1320

QY 1321 GYNGERADPVSGVTHLNGYRSYDPTLMRPHTPDLSLSPFGAGGINPYSYCLGDPINRSDP 1380  
DB 1321 GYNGERADPVSGVTHLNGYRSYDPTLMRPHTPDLSLSPFGAGGINPYSYCLGDPINRSDP 1380

QY 1381 SGHLSWAQWTGIGMGIAGLLTIATGMAIAAAGGIAAAIASTSTTALAFGALSVTSDIT 1440  
DB 1381 SGHLSWAQWTGIGMGIAGLLTIATGMAIAAAGGIAAAIASTSTTALAFGALSVTSDIT 1440

QY 1441 SIVSGALDASPKASSILGWVSMGMAAGLAESAIGKSTKLATHLGAFAEDGENALLKST 1500  
DB 1441 SIVSGALDASPKASSILGWVSMGMAAGLAESAIGKSTKLATHLGAFAEDGENALLKST 1500

QY 1501 SSSRIKWGVTRSLDREIVRNEEQVVKDHSRGYTDNFMKGEOAILVHGDGKDFLYHTE 1560  
DB 1501 SSSRIKWGVTRSLDREIVRNEEQVVKDHSRGYTDNFMKGEOAILVHGDGKDFLYHTE 1560

QY 1561 GNKNGKGPYTRHTPEQLVDYLDKNNIVDLTQGGDKPVLHLLSCYKSSGAADKWKAYINR 1620  
DB 1561 GNKNGKGPYTRHTPEQLVDYLDKNNIVDLTQGGDKPVLHLLSCYKSSGAADKWKAYINR 1620

QY 1621 PVIAYSNKPTISQGLARIERKDFFLKSTYHSYDPRKIIILGRTEKTVKPKTRP 1673  
DB 1621 PVIAYSNKPTISQGLARIERKDFFLKSTYHSYDPRKIIILGRTEKTVKPKTRP 1673

## RESULT 2

AAY33727

ID AAY33727 standard; protein; 1584 AA.

XX

AC AAY33727;

XX

DT 09-NOV-1999 (first entry)

XX

DE Photorhabdus luminescens 176 kD insecticidal toxin.

XX

KW Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.

XX

OS Photorhabdus luminescens.

XX

FN WO9942589-A2.

XX

PD 26-AUG-1999.

XX

PF 18-FEB-1999; 99WO-EP001015.

XX

PR 20-FEB-1998; 98US-00027080.

XX

PR 20-JAN-1999; 99US-0116439P.





Db 1007 LGVSPLOQTYSYLLPMWLLHAEVEVGQAQSAWD-RWPSGLR-EETHDIRSGKKKAHYR 1064  
 Qy 988 WUURLENGVDTUTGIQKISRTD---H-GRVQIKDSSIKTTFLYDLDNRHIGSOVTDL 1043  
 Db 1065 YSLTGNLEGADIDGAHARSYETAHAHVKGKLEIAADAAVTVTLAYDGLQRLCSWTARD- 1123  
 Qy 1044 ATGHMLTTVEFGLNREIGRKLCDSSGHTLDIQOSWLKTOQLANRIVKLNGLVQLQTEQY 1103  
 Db 1124 GRGHAAITLLEFSLGRKTRTAAESAEATLSQWYPNGQLHQRKSEGGKPFCDTEF 1183  
 Qy 1104 SYDSRNLNOKCDGAECPDQKYGHSIVQNFYDIYGNITACHTTFADCTEDHATFKA 1163  
 Db 1184 VYDARNLDYASGGLPKDAIGNAIRGQKEFFDAFNIRKCTTVLDGSENVGEYLF 1243  
 Qy 1164 NPTDPCOLTEVHTHPD--MPDNIRLYDKAGRVINITDNHNTENTYDTTLGRL---QN 1218  
 Db 1244 NPADPCQLTKVNSALDKGYPPAIELKYDAQGLER--DEAG--RRLSYDALGLRLARVEG 1299  
 Qy 1219 GQGSV-YGYDPLNRLYSQKTD--LCELYYRETMLVNE--VRNG-----EMIRLLRT 1266  
 Db 1300 GGSASGYGDAHDRLVCQVETSGMDHRLYFRANRLVNEWMTRSGQAPGADDRLVYA 1359  
 Qy 1267 GETIIAQOR---ASKVLLTGTDSQOSVILTSDQNLSQEAAYSAYGKHST 1313  
 Db 1360 AGSCAAQVNEGDSVAALMGTDKGSIVSQAEQQAQKHYATPYG-HOSS 1409

RESULT 4

ABM67433  
 ID ABM67433 standard; protein; 1787 AA.  
 AC ABM67433;

20-NOV-2003 (first entry)  
 DE Photorhabdus luminescens protein sequence #530.

KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.

OS Photorhabdus luminescens.

XX WO200294867-A2.

XX 28-NOV-2002.

XX 07-FEB-2002; 2002WO-IB003040.

XX 07-FEB-2001; 2001FR-00001659.

XX (INSP ) INST PASTEUR.

XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunst F, Danchin A;  
 XX Buchrieser C;

XX WPI; 2003-148459/14.

XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.

PS Claim 2; SEQ ID NO 530; 1205pp; French.

XX The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that

CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins

XX SQ Sequence 1787 AA;

Query Match 17.8%; Score 1583; DB 6; Length 1787;

Best Local Similarity 28.2%; Pred. No. 3.6e-101;

Matches 511; Conservative 280; Mismatches 727; Indels 294; Gaps 65;

Qy 22 TOANNFTSAVSGVDPRGTGLYNIQITLGHIVGNLGPLTLPITLSYSPLNKTDIGFGIGF 81  
 Db 30 SQATNFTGAMGSGVDPRGTGLTTPQIPITQLGNDLWGPDLAITLVNPLTYLNTGYGLGF 89  
 Qy 82 NFGLSVYDKNSLLSLSTGENTYKVTETDKTV-----LQKKKLNLRFEKDKLNCYRI 135  
 Db 90 SDNFTRYDTQTQVLTATGEIYHRVEKANEVVDGQAWTFPHAKPAHFVKKE--KDAFWV 147  
 Qy 136 IHKSGDIEVLTCFNNAFDPKVPKLLMPAGHAIYIDNFEATQP-----RLNRIYDLDLG 191  
 Db 148 LYKTSREKLTOL-DRANPVAVVEIYAPSGHKLCKVKNVSFVNNHNYNQLMEVCDAE- 205  
 Qy 192 HDIPLNLLEYQGLIKTILTPPGQKEGYRTERFLNRLNLSIHNFSLGNENPLTWSFGYT 251  
 Db 206 ---TLKADLATTEKIEFTVWFGSPESYTVTLNMTNDLLQTVISAS-----ELTWHLEYE 257  
 Qy 252 PIGKNGILGQWITTSMTAPGGLKETVYNNNGGHHPP----- 288  
 Db 258 TEGAHKNI---LTKVTPSGLTKEVYV--HETGHTLPTPKCIQYYPNAGPGIIRQDPKS 312  
 Qy 289 -QSANLPVLPYVTLMKQVPGAGQAPAIQAEYSYTSNHYVGGSGNGIWN-----NKLNLNLYL 343  
 Db 313 STTATVEHFYVYVTHHIIAGSGSPQVIRYVSPENFLGQKNMKDPILPLQDQDAYLA 372  
 Qy 344 MTEYNYGSTEGRYKDKEGHDOIVRIERTYNNYHLLTSECK-----QONGYIQTETAY 397  
 Db 373 NSEYKYTSTEVEYNEKR-----YCIHREYNKFLHLLVSETEVETVTPSPQKLKELIKY 427  
 Qy 398 YAIIGHNFD--SQSQFOLPKTKTETWRSADNS-----YRSEITETTFDESGNPLTKVIK 450  
 Db 428 YADVGSFDDNKQPPQFLMPNVTETIWHNPESASTTQRRKETTQWEYNAQCNLISMTLP 487  
 Qy 451 DKTKOKIISPSTHWEYYPAGEV---DNCPEPYGTFRFVKKI--TQTPYD-SEFKDDDPK 505  
 Db 488 DNTTKT-----TYAPDGEETTDTHCPAEPNPFERFIKEIAVEAPSPLTKITILR 540  
 Qy 506 FIQYRYSLIGSOS-----HVTLKIEERHYS-----ATQL-----LNSTLFQYNTD 545  
 Db 541 KTYNYKSYDYSPPNNQASVVKSWLPLSTHYSRRDCCADHLEKVKINTVSVFNTQ 600  
 Qy 546 KSELGRLLKQTECTKGNG-KTYSVYVHKFTYTKQDDTLQQSHSITTHDNF-----TIHRS 599  
 Db 601 NAFLNGRVEQRNSYLFENGNGQTSYTEDYSWSENKNGASCICKTKTGGCKGCTPPVSHE 660  
 Qy 600 QVRSYRTGRLSDDTDTKDIIVTQMSYDKLGRLLTRTLNSGTPYANTLTVDYELNLDNR 659  
 Db 661 QYWSRSTGRLLFPQDAQDNNTVFDYDTTGLHLSSTINADTAYERTVKYAWSNKR----- 716  
 Qy 660 PPFTVTTDNGNQLRNEFDGAGRHSVSOCLKSDGDKGFYTHHTQYDQGRHHTSTYSD 719  
 Db 717 --VTVQTDIHENRIVITEMDGLGRPELKKYSYPAGYQGFQDMERYQYNPLQGLQCAISCD 774  
 Qy 720 YLTNGRQQTDPDKVHLSMSKSYDNGWQIANTHWSYGVSEKITTVDPTITLTATKQLOSNN 779





180 PRNRIYDDLDGHDIPLLNLEYOGLIKTLTL-FPGQEGYRTELRL-NRQLNSIHNS 237  
117 ----EYFDLIYNSRFALSEIKRTWKTYLKLNSGNN-----CTSVETPDNNISAKIAFD 169  
238 LGNENPLTWGFTYPIGK-----NGILQWITSMTAPGLKETVYNNNOG 284  
170 YRNDYLITVTPYDASPIDSARFKMTYQTLKGF-PVISAFRTPTGYVELVSYKEN--G 226  
285 HHFPOQANLPVLYVTLMKVPGAGOPAQAEYSYTS-HNYVGGGNGI-WNNKLNLYG 342  
227 HKYTDTESIPYAAALTIQ---PGNGOPAIKSYEYSVHNFGLGYSGRTSFSDSSQDNLYL 283  
343 LMTEYNVGPESRRYKDKGHDQIVRIERTYNNYHLLTSECKQONGVIOITETAYALIG 402  
284 VTGKITYSSIE-----RVLNGQNVISVTERVDFKHLMTKEAKTODNKRITETIYNEDEPS 339  
403 HNFDSQSPQLPK-----TKTETWRSADNSYRSEITETTFDESIGNPLTKVKDKKTKQ 456  
340 KSFSEQENLQOPSHVLTTRYDLOQ-----NTSREESVNIKSDDWGNTLL-ITETSGIQK 393  
457 IISPSTHWEYPPAGVENCPEPYGFTFPVKKIIQTPYDSEPKDPEKFIQYRYSLIGS 516  
394 -----EYVYYPVNGEENCPADPLGFSRFLKSVTKGSPDAAQSVANKRVTSYTYQKLP 447  
517 QSHVTLKIBERRHYSATQLLNSTL---FOY---NTDKSELGRLLKQTECKGKNGKTYSVV 570  
448 FTGAYVK--EYVSKASETIDSKIVRTFNYSNPTNKS--GSLAKITSVMNQ-----QTV 499  
571 HKFTYTKQDDTLQOQSHSITTHDNFTIHRSOVRSRYTGRLLFSDTDDTKDITVOMSYDKLGR 630  
500 TTFKYEYSSEMTNSTVTGFDGTHMESKNVTSIYTHRQLRKVDVNVHVTIDQSDLSGR 559  
631 LRTLNSGTPYANTLYDELANNLQDNRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLK 690  
560 IGQIIPDGTIKTKRSYIYQPGDENDFWP-VMEIDSGIRKTHYDGMGRICISIEQ 618  
691 DSDG-----DGKFVTHITQYDEQGRHHTSYSDYLTN-----GROQTPDKVHLSM 737  
619 DDDGVWGTSGIYQGYRKLARQYDVLQGVKBEISNDLWLDLSANPLRLTTP--LVTTK 676  
738 SKSYDNWQIANTHMSGVSEKITVDPITLTATK-----QLQSNNSNVQTGKEVITYTP 791  
677 TYQYDGNWGRNYSYSDGRIELEIHPITRTITQGVKGLMLNIQNNF-----725  
792 SQOPIQITLDFEAGHLQSCHTLTRDGDWRVRKETDAIGCTIYQYDYNVRVIOITLPDGT 851  
726 -EQPASIKVYVPGAIYSTYTYDGFGRVITETDAEGYATQIEYDLFDRIVKTLTLPDT 784  
852 IVNRKYAPSTDTLITDIRVNGISLGOQTFDGLSRLTQSDQGRVWAYTYSAGNDQCPST 911  
785 ILESAYASFHEELISALANVNGTQLGSLVYDGLGRVTRDTVGGKRTXYLYGSQGDK-PIQ 843  
912 VITP--DQGFHYQYQPELDDAVLOVASNEITQOFSYNPVTGALLKAVAGQSLTPI-YY 968  
844 SVTPAHKNQONIDLY--ALGSVMSKFTTETSQNFYSYQYKTGALLSA-TEGVSSQSNYSF 900  
969 PSGLKXENIN-DMKXMS---YLWTLRGLNGVYDGLTGTIKISRDTGRVTOIKDSSIK 1024  
901 PSGLVQHEFSRNRKPISSGDYRYTMSGLIQRKDSFAHDHVYSYDAEGLRVKTESSQY 960  
1025 TTLNYDDLNRHIGSQVDTLATGHMLTTTVFDFGLNREIGRKLDCSSGHTLDIQOQSLTKQ 1084  
961 ATFEYDNVGRLLITTTTKDITSLSQLATKIEYDFDREIKRSLISDFSIOV-ITLSYTKN 1019  
1085 QLANRIVKLVGLVQRTQYQSYDSRNLNQYKDGAECPDTKYGHSIVTONFTYDIYGNIT 1144  
1020 QISQRTSIDGVVMKRYERYQYDSNQRLSQYQCEGEQSPVDHTGRVLSQOIYHYDOWGNIK 1079  
1145 ACHTTADGTEHDATKEANPTDPCOLTEVHHHTHPDMPDNIRLKYDKAGRINIT-DNKG 1203  
1080 RLNTYRDGKET-VDXHFSQ-ADPTQLIRI-----TSDKQOIELUSDANG---NLTRDEKG 1130  
1204 NTENFTYDTLGRL---QNGQGSV---YGYDPLNRLVSQ--KTDTLDCLEYRETMLVNEV 1255

1131 QT--LIYDQNNRLVQKSKGNLVCQYQYDALNKLTAQVLANGTVNRQ-YYASGNVANVQ 1187  
1256 RNGEMIRLL-----RTGETIIAQORASKVLLTGTDSQQSVILTSQKQNLISQBA 1303  
1188 LGDETITLWSSDKORLGHQSTKNGESVYQY-----GTDHNSVVIASQENELMALS 1239  
1304 YSAYGKHKSSTANDASILYNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPTGAGG 1363  
1240 YTPYGFRLI-----SSLPLGLNGAQVDPVTGTYFLNGYRVFNFPVLMRFHSPDSWSPFGRGG 1296  
1364 IMPYSVCLGDPINRSDPSGHLQWQAWTGIGMGIAGLLTIAT--GGMAIA---AAAGGIA 1417  
1297 VNPYITCOQDPINRIDLNGHLSAGGILGIVLGAIGIIVGVSLGAGAAISAGLIAAGGAL 1356  
1418 AAIASSTTTALAFGALSVTSDITSIVSGALEDASPKASSILGWVSMGMAAGLAESA1-- 1475  
1357 GAIASTSAFATVATVIGLAADSIGIASAALSEKDPKTAGILNWIWISGLVLSFGISAITF 1416  
1476 -----KGGTKLATH--LGAF-ABDGENALLKSTSESSRIKMGVTRS 1513  
1417 TSSLIKARSQGVASTSVIGSVPIEFGEIA-----SRSSR-RWDIALS 1459  
RESULT 6  
ADG73113  
ID ADG73113 standard; protein; 982 AA.  
XX AC ADG73113;  
XX DT 11-MAR-2004 (first entry)  
XX Pseudomonas syringae pv. tomato DC3000 Avr/Hop protein #53.  
XX Avr: Hop; transgenic plant; disease resistance; cancer; bacteria;  
XX metabolic pathway; eukaryotic cell death; programmed cell death;  
XX cytostatic.  
XX Pseudomonas syringae; pv. tomato str. DC3000.  
XX US2003204868-A1.  
XX PD 30-OCT-2003.  
XX 12-FEB-2003; 2003US-00365742.  
XX PF 12-FEB-2002; 2002US-0356408P.  
XX PR 10-MAY-2002; 2002US-0380185P.  
XX (COLL/) COLLIER A.  
XX PA (ALFA/) ALFANO J R.  
XX PA (CART/) CARTINHOOR S W.  
XX PA (SCHN/) SCHNEIDER D J.  
XX (TANG/) TANG X.  
XX Collmer A, Alfano JR, Cartinhour SW, Schneider DJ, Tang X;  
XX WPI: 2003-875735/81.  
XX DR N-PSDB; ADG73112.  
XX  
XX New nucleic acid, useful in imparting disease resistance to a plant or in  
XX preparing a composition for treating cancer.  
XX  
XX Claim 15; SEQ ID NO 106; 209pp; English.  
XX  
XX The present invention relates to the isolation of Pseudomonas syringae  
XX pv. tomato DC3000 Avr/Hop proteins, and the polynucleotide sequences  
XX encoding them. Also disclosed are expression vectors, host cells, and  
XX transgenic plants comprising polynucleotide sequences of the invention.  
XX The polynucleotide and polypeptide sequences are useful in imparting  
XX disease resistance to a plant or in preparing a composition for treating  
XX cancer. The sequences may also be used to make a plant hypersusceptible  
XX to colonisation by nonpathogenic bacteria, modify a metabolic pathway in





Db 294 -----TTETD-----PFGNKITYSFDEHF-- 313  
QY 432 EITET-----TPDESNPLTKVKKKTKOKIISPSHWHYPPAGEVNDNCPPEYV 482  
Db 314 ELNETNQLGVTISYADENGRIISVNNENSKTKLA-----YDANGNIK-TTNPLG 364  
QY 483 FTRFVKKIIQTPYDSEFKDDPEKFIQVRYSLIGSQSHVTLKIBERHYSATQLINS----- 537  
Db 365 YSK-----SMTYDS-----KNLISQDELHRKTSFEYDDNSNLKSIDALG 406  
QY 538 --TLFQYNTDKSELGRLLKQTECK-----GENK-----TYSVVHKFT 574  
Db 407 HETVFSYD-----KYGVIGETDSNKKKTATFSYNNNGDQITITDANGKTSFTYDTVGRVT 462  
QY 575 -----YTKQDDTLQQSHSITTHDNFTIHRSOVRSYTGRL-----FSDTDTKDI 618  
Db 463 TKTDAGKNRYTFQYDALDNLISITD-----PMGQTTSTNTMLLEIKVLQMLKVDXQSI 516  
QY 619 VTQMSYDKLGRLLTRTLNSGT-----PYA-----NTLTYDYELNNLQDDNRPFPVIT 665  
Db 517 LTTLXISLXKRMQWVEXLGTNNMPLAIWPPRTQKGHKTSYDYDPLNRQ-----VS 568  
QY 666 TTDVNGNQLANEPDAGRHVS-----QCLKDSGDGKFTYIHTQQQYDE 708  
Db 569 VTNALGKTRNKYDAIGNKISITNAYGKSTRYSYNSLNQLVKVTNAMGR---VVRVNYDA 625  
QY 709 QGRHHTSTYSD-----YLTNGRQOTDPDKVHLMSKSYDNWQIANTHWSYGVSEKIT 761  
Db 626 VGNLSTTBENGKINYGDSLNQRQSVTDALRKNKYDAVGNKISITNAYGKSTRYS 685  
QY 762 VDPITLTATQLOSNNVQTKGVITYTPSOQPIQITLFDGAGHLQSCHTLTRDGDWRV 821  
Db 686 Y-----NSLN-----QLVKVTDAMGGVVRVNYDAVGNLIS----- 715  
QY 822 RKETDAIGQTIQYDNYNRVQITLPGDTIYNRKYAPFSTDTLITDRVNGISIGQOTF 881  
Db 716 --TTDANGRTKTYGDSLNQRQSVITNALGKTRNKYDAVGNKISITDANWR---LTKYSY 770  
QY 882 DGLSRLTQSDG-GRVWATYSA-GNDQCPSTVITPD--GQFIHYQY-----QPELDNAV 932  
Db 771 DSLNRLVKVTDAMGGVVRVITYDAVN-----LISTDAGHKTDYGYSLDRQVSTIPL 825  
QY 933 LQVASNEITQFSPYNPVTGALLKAVAAGOSLPIIYPSGRGRKVENINDMKMSYLMWLRG 992  
Db 826 GRTARNK-----YDAVGNKISSTDEGKTSYGYDVNLRLKVSYPDDQKVS----- 873  
QY 993 LENGYTDLTGTTQKISRDRTHGRVTOIKDSIITLNYDDLNRHI-----GSQV--TDLA 1044  
Db 874 -----NYDAVGNRLTWKDSHGTTAYKYDKLNRLSLVLPNDGQKVSYYTNK 918  
QY 1045 TGHML-----TTTVEFDGLNREIGRKLCDSSGHTLD---IQSWLKTQQLANRIVKL 1093  
Db 919 VGNRVKMTYDPDKTTSYDAVNLIG--VIDSDGHITSYAKNGNLKMTWPNGV--- 973  
QY 1094 NGVLRQTEQYSDSRNRLNQYKDCGAECPDKYGHISIVTQNFPIYDIGNITACHTTFADG 1153  
Db 974 -----KTE-YSYDKANRL-----VELINKNTQVWSSYKYTLDAAGNRLKVDQSLAEG 1020  
QY 1154 TED-----HATFEFNPDPQCLTEVHHTHPDPDN--IRLKVDKAGRVINIT- 1199  
Db 1021 VESGSELKESQLLTTTYGY---DKLYRLTKV-----DYPNKTYSYKYSNGNISMSTT 1072  
QY 1200 --DNHGNTEFNFTYDLTGLRQNGQSVGYDPLNRLVSOKTDTLDCELY-YRETMLVNEVR 1256  
Db 1073 NYVDGIGSTISYKDAADQLQSGNYSYDKKGNLKKRVNSTQPMYSYDE---ANRLK 1129  
QY 1257 N-----GEMIRLLRT-----GETIIAQQRASKVLLTGTDSQOSVILT 1293  
Db 1130 NYSEFVSNTNTPKSSYNFYDGDGNRIIKTTINGE---NAQSTKYVLDINSALPOVLTE 1185  
QY 1294 SKQNL-----SQEA-----YSAYGKHKSTANDASTIL----- 1320  
Db 1186 SDTKNTTCYTGTDLISMNTSENABYYHYDGLGSVRSLSDSKGIKNKYLYADFGQVQK 1245

QY 1321 -----GYNGERADPVSGVTHLNGYRSYVDPTLMRPHPTDLSL--PFGAGGINPYSY 1369  
Db 1246 EIGTVDNBFRFTGEQMDDETGLIYLARY--YDPSVGRFITKDVIKGRRTVTQTSINRYVY 1303  
QY 1370 CLGDPINRSDPSGH 1383  
Db 1304 TTNNFVNLVDLTGY 1317  
RESULT 9  
ABG31849  
ID ABG31849 standard; protein; 2334 AA.  
XX ABG31849;  
AC ABG31849;  
XX 05-NOV-2002 (first entry)  
XX Human kinase, MEK1.  
XX Extracellular signal regulated kinase; hyperalgesia; surgery;  
XX opiod withdrawal; pain sensitisation; analgesic; chronic pain; ERK;  
XX MEK1; human; enzyme.  
XX Homo sapiens.  
XX WO200258687-A2.  
XX 01-AUG-2002.  
XX 25-JAN-2002; 2002WO-US002128.  
XX 25-JAN-2001; 2001US-0264336P.  
XX (TEXA ) UNIV TEXAS SYSTEM.  
XX Gutstein HB;  
XX WPI; 2002-608414/65.  
XX N-PSDB; ABK90804.  
XX Reducing or reversing tolerance, physical dependence, hyperalgesia,  
XX withdrawal symptoms, or pain sensitization in patients on analgesics for  
XX chronic pain, comprises inhibition of the extracellular signal-regulated  
XX kinase (ERK).  
XX Disclosure; Page 156-161; 163pp; English.  
XX The invention relates to a method of reducing or reversing tolerance,  
XX reducing the risk of physical dependence or hyperalgesia, reducing the  
XX symptoms of opiod withdrawal or inhibiting pain sensitisation in a  
XX patient taking analgesics. The method comprises administering an  
XX analgesic and an extracellular signal-regulated kinase (ERK) inhibitor  
XX comprised in a formulation to reduce or reverse tolerance, risk of  
XX physical dependence, hyperalgesia, symptoms of opiod withdrawal, or  
XX inhibiting pain sensitisation in patients taking analgesics for chronic  
XX pain or those undergoing surgery. The present sequence represents the  
XX amino acid sequence of human MEK1 (not defined)  
SQ Sequence 2334 AA;  
Query Match 4.9%; Score 432; DB 5; Length 2334;  
Best Local Similarity 20.4%; Pred. No. 4.5e-20;  
Matches 358; Conservative 226; Mismatches 613; Indels 560; Gaps 80;  
QY 44 IQITLGHIVG-NGNL-----GPTLPLTLYSPLNKTDIGFGIGNFGLSVYDR 90  
Db 896 IDIPSGQLNGATGNVIVNEEDLSIDGRGPGGLGSLRTVNSLSDSDHLFGQW----YADA 950  
QY 91 KNSLLSLSTGENYKVIETDKTVLKQKKLDN-----LRPEKDLKENCYRIHKSQDIEVL 145  
Db 951 ETSVISTDGGAMY--IDEDATTHRFTKADGTYPPTGVVYLELTETADQFILTKD---- 1004

QY 146 TGFNNAFDLKVPKLL-----NPAGHAIYIDWN-----FEATQRLNRIYDLDGH-- 192  
Db 1005 ---QTNAYFNKGKGLQKVVDGHNNAVTYNDKNQLTAITDASGRKLTFTYDE--NGHVT 1060  
QY 193 -----DIPILNLEY--QGLI-----KTLTLFFG 214  
Db 1061 SITGPKKKVTSYENDLLKKVTDGTVTSYDSDSEGRVLKQYSANSTEAQPVFTEY-- 1118  
QY 215 QKEGYTELRFLNRQLNSIHNF-----LGNENPLTWSFGVTPGKNGILQOWTSM 267  
Db 1119 QYSGHLEKAINAKETVYVSYDADKKTLLMTOPNGRKQVGYNEAGNP-----IQVID 1172  
QY 268 APCGLKETVN--YNNN-----QCHFPQSANLPVLPVYTLMKQVPGAG 309  
Db 1173 DAELGKITNTKTEGNNVVEDPNDVGTGKATESYOYDKGN-----VTSKDAYGT- 1225  
QY 310 QPAIQAEYSYTSNHYVGGSGNGIWNKNLONLYGLMTEYNYGSPESRRYDKKEGHDQIVRI 369  
Db 1226 -----ETYEYKNDV-----TKMKDTESVNTDIAVDGLDAYSETDQSGKSSAAV 1271  
QY 370 ERTYNNYHLLTSECKQONGVIOTETAYYAIIGHNFDSPQSPQOLPKTKTETWRSADNSY 429  
Db 1272 YDKYGNQIOSKOLSASTNLK-----DGSPEAQSGWNLTASKD----- 1311  
QY 430 RSEITETTFDESNGPLTKVKKTKOKIIPS-----THWEYPPAGEVNCPPPYGFTR 485  
Db 1312 RRKIS-VIADKSG-----VLSSGKALEVLSQSTAGTDHGYSSATQVLEPNTTYTSLG 1365  
QY 486 FVKKII---QTPYDSBFKDDPEKFIQY---RYSLIGSQSHVTLKIEERHYSATOLLNS-- 537  
Db 1366 KIKTDLAKSRAPNIDLRKDKQRIQWINEVSALAGKNDWT-----KROITFTTPANACK 1421  
QY 538 ---TLFQYNTKSELGRL-LKQECTKGNGKTYSVVHKFTYTKQDDTLQOS--HSITTHD 592  
Db 1422 AVVMEVDHKDKGKAFDEQLEKGEVSSSYNPVQSSPTSATENNWNVSASVDSEE 1481  
QY 593 NFIHRSQVRSYRTGBLPSDSTDQDIQVOMSYDKLGRLLTRT-----L 635  
Db 1482 GFNDVDSLKAARTSASQAGSVTKQTVVLGQSKNDKPVYLTLTGMSKASSVKFTDEKDYSL 1541  
QY 636 NSGTPYANTLTVDYEL--NNLQDDNRPPFVI----- 664  
Db 1542 QANVYADSGTGIYNAKFPSTGQEWNRAAVVPKTPINKVDISILFQKSATGTWFFDI 1601  
QY 665 -----TTTVDNGNOLRNEFDGAGRHSQCLXSDGDKFYTIHTQQYDEQGRHHT 714  
Db 1602 RLIEGSLTKSTYDSNGNVVTKDEELGYATS---TDYDETK---KTSETDAKGEKT 1654  
QY 715 STY---SDYLTNGRQOTDPKVLHLSMSKSYDNWG-QIANT-----HWSYGVSEKI-- 760  
Db 1655 YTYDQADQLTNMTLSNGTSLH-----SYDKEGNEVSKTIRAGADQTYKFEYDVMGKLVK 1709  
QY 761 TVDPITLTATKQLOSNNVQT---GKEVTYTPSQQPIQITLPDEAGHLQSCHTLTRD 816  
Db 1710 TTDPLGNVLASEDANSNLTKTISPNGNEV-----SLSYD 1744  
QY 817 GMDVRKETEIDAQCCTIYOYDYNRVIIQITLPDGTIVNRKYAPFSDTLITDIRVNGISL 876  
Db 1745 GTDRVKSYSYNGTEKIFYTDKNGN-----ETSVNKEQN-----TT 1781  
QY 877 GQOTFGLSRLQSQDGGRWMAITYSAGNDQCPSTVITPDGQFIHYQVPELDDAVLQVA 936  
Db 1782 KKRTFDPKNRLTELTRDQSGQWTWYPSDSKLKTF-----SWIH-----G 1821  
QY 937 SNEITQOFSYNPVTGALLKAAVAGQSLTIPIYPSGLKMNINDMKMSYLTLRGLENG 996  
Db 1822 DQKGTNQFTYN-----KLDQMIEMKDSSTSSYFDYDEN- 1854  
QY 997 YTDLTGTIOKISRDTGRVTOIKDSSIKTLLNYDDLNRHIGSOVTDLATGHMLTTTVFED 1056  
Db 1855 -----GNVO-----TFTGNGGGHSFYSDERNLVSLHIGDKNGGDLITSEY- 1898  
QY 1057 GLNREIGRKLCDSSGHTLDIQOSWLKTOQLANRIVKLNGLVQRTQEQSYDSRNLNOYKC 1116

Db 1899 -----DANGNRRTTINS-----ASGKV-----QYEGKLNQLVK--- 1927  
QY 1117 DGAECPTDKYGHISIVTQNFYDIYGNITACHITTFADGTED--HATEKFANPTDPCQLTEV 1174  
Db 1928 -----ETHEDGTVI---EYTYDGFGRKTV--TIKDGSSKTVNASFNWN-----QLTKV 1973  
QY 1175 H-----HTHP-DMPDNI-----RLKYDKAGRVNITDNGH 1203  
Db 1974 NDESISYDKNGNRTSDGKFTYTWDAEDNLTAVTKKGEDKPFATYKYDEKGNRIQKTWN-G 2032  
QY 1204 NTEFNFTDLGLQNGQSGVGYDPLNRLVSKTDTLDCELY-YRETMLVNEVRNGEMIR 1262  
Db 2033 KVTNIFYDG-----DSLNVLYETDADNNVTKSYTGD-----SGQLLS 2070  
QY 1263 LLRTGETIITAQORASKVLLTGTDSQOSVILTSDKQNLQSOEAYSAYGK-HKSTANDA---S 1318  
Db 2071 YTEGKKYFYHNAHCDIIAISDSCTKTV-----AKYQYDAGNPTKTEASDEVKDN 2122  
QY 1319 ILGNGERADPVSGVTHLNGYRSYDPTLMRPH--PDSLSPPGAGGINPYSYCLGDPIN 1376  
Db 2123 RYRYAGYQYDEETGLYLYLMARY--YEPRNGVFLSLDPPDPSGDSLDQNGAYAYGNPNVM 2180  
QY 1377 RSDPSGHLSSQAWTGIGMGIAGLLLTIATGGMALAAAG---GIAAAIATSTTALAFGA 1432  
Db 2181 NYDPPDGH--W-VMLVNVNAGFA-----AYDGYKAYKSGKMGKGAWAAASNFGPKIFKG 2231  
QY 1433 LSVTSDITSIVSGALEDASPKASSILGWVSMGW---GAAGLAESAIGKGTKLATHLGA 1488  
Db 2232 ASRAYKFTK-----KAVKITGTRHGLNQSIGRNG-----GRGVNLRKALNA- 2273  
QY 1489 AEDGENALLKSTESSRIKMGVTRSLDRE--IVRNEEGQVIKDHRSRGYTDNFMKGEOAI 1546  
Db 2274 -----VRSPPKVIKQPNGATKYVKKATVVLNKRGVITAYG-----SS 2312  
QY 1547 LVHGDKDGLYHTEGKN 1563  
Db 2313 RAKSGKHVFHTHKGKNK 2329  
RESULT 10  
ABU18641  
ID ABU18641 standard; protein; 2234 AA.  
XX AC ABU18641;  
XX AC  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #4168.  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX Bacillus anthracis.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-02926/02.  
DR N-PSDB; ACA22511.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 46565; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 2234 AA;

Query Match 4.3%; Score 386; DB 6; Length 2234;

Best Local Similarity 18.7%; Pred. No. 7.2e-17;

Matches 339; Conservative 236; Mismatches 633; Indels 608; Gaps 76;

QY 21 FTQANNFTSAVSGVDPR-----TGLNIQITLGHV-----NGNL----- 57  
DB 748 YSKVGHAESEIHFVPRFSQSGFIMFDYWASIPVLNGKVNATNGNFMSEKDIITLSG 807  
QY 58 -GFTLPLTSLYSLNKTIDGFGIFNGLSV---YDRKNSLLSL-----TGEN- 102  
DB 808 RGPDSVVERTYNSQSKVGLFGTGWSSGLEERWADGNGNLLIISTDGNITFTTGDNK 867  
QY 103 -----YKVIETDVTVKLQKKLNLNLFPEKDLKENCYRIHKGSDIEVL 145  
DB 868 YOAPTGTIVLEIKQVSGVEIKDKDQTV-----TFYKSDAQGR 905  
QY 146 TGFNNNAFLKVP-----KLLNPAGHAIYIDWNFEATQPLNRIYDDLDGHDPL 196  
DB 906 IBYTKDKYNTTYEYDGNASRLSKVKNASGKELVL--QIDGNKKKAARVI-----GPDNKT 959  
QY 197 LNLLEYQGLIKITLTPFGQ--KEGYRTLELFLNRLQNSIHNFLGNENPLTWSFGY--- 250  
DB 960 ITFNYDGLLVSTTPEGKVKYGYD-----NGVLTSDYDPQHTDAKPKYKTSYAYENDR 1013  
QY 251 -----TPGKNGIL-----GQWITSMTAPGGLKETVYNNNGHHFPOSANLPVLPYVTL 301  
DB 1014 LVKVTDPGLKATTLAYNTGSKSEVTLTPNPKRKTVTYND-----AGNPV----- 1057  
QY 302 MKQVPGAGOPAIQAEYSYVSHNVGG-----GSGNWNKLDNLYGLMTEYN- 348  
DB 1058 -KTVEDVGRNLNLTTSYEYANNLVKTPPKQGTETATYDNGNGVNTSDMEGTEKFEYNK 1116

QY 349 -YGSTESRRYKDK-----GHDOI VRIERTYNNYHLLTSECKQONGVYQIOTETAYVAI 400  
DB 1117 DNGIIKATNEDRKTTTAYVAGNTEVSQDQAN-----TSSVIHDOYGNPIETSKELS 1171  
QY 401 IGHNFDSQFQOLPKTKTETWRSADNSYRSEITETTFDESGNPLTKVIKDKKTKIISP 460  
DB 1172 AGGNLIQFNS-FEM--NGTEKWKVVDVNNSGSIS-----KDATPAPGLGGSLSKITTK 1223  
QY 461 STH--WEYYPAGEVDNCPPEPYGTRFVKII-----QTFYDSFCKDDP 503  
DB 1224 ATNNDWGYIAAIQEVLTLEPNTTYTLSGMVKTDLVNGAAFFNVQSLNENGAGIDGGWHDR 1283  
QY 504 EKFIQYRSLGSSQSHVTKLKEERHYSATQLLNSTLFOYNTDKSLGRL-LKQETCKGE 562  
DB 1284 HNKVQGTSDWNRQ--VTFKTEQ-----TRAKVIYQVENGSGSATSGAWFDKIQLEKE 1337  
QY 563 NGKYSVVHKPYTKQ--DDTLQO-SHSITTH-----DNFTIHRSOVRSRYTRLFS 611  
DB 1338 VSSSEFVLNSSFEEENWDPGFVPQWVRSCSQHCERNVDSDSFTGHSSIVMER-----S 1391  
QY 612 DTDTKDVIYQMSYDKLGRLLTFLTNSGTPYANTLYDYELANLQDDNRPFPVITTYDVG 671  
DB 1392 EYGPNDI-----GYRNRVILNQKAEVTVTLTAMSKSENVDN----- 1428  
QY 672 NQLRNEFDGAGRHSQCLKSDGDKFYTIHTQOYDEQGRHHTSYSDYLTNGROQTDp- 730  
DB 1429 -----APDKLSKYAVLAETYYQDG-----TVVNYTTSFSGTNDW 1464  
QY 731 -----DKVHLSMSKSYDNMGQIANTHWSYGVSEKITVDPITLTPATKQLOSNS 777  
DB 1465 NRSAAVIPAKKPIQKIEIFLLFRKNKG-----KWFDDIRLEGNALIKNE 1511  
QY 778 NNVTGKEVYTTTSQOPIQITLDEAGHLSCHTLFRDGDWRVRKETDAILGCTIYQD 837  
DB 1512 YD-NDGNVAVY-----DEEGOK---NTFTYDASGNKSEKSETDGNKTKLYDYN 1555  
QY 838 NYNRVIQITLPGDITVNRKYAPFSTDTLITDIRVNGISLGQOTPDGLSRLTSQDGGRW 897  
DB 1556 KDNLLTKVLKNGTSVNYRD-----RNGKDSFT-----FERDQNGQVTK 1670  
QY 898 AYTSAGNQCPSVTITPDGQFIHYQOPELDDAVLQVASNEITQQFSYNPVTGALLKAV 957  
DB 1577 ---HNGNTTEKSYVMFGGKTQTHKYEYDNDKNTVYIDALNRRIEN-TYDENANKIKTKM 1631  
QY 958 AEGQSLTPIYYPGRKLKMNENDMKMSYLTWLRGLNGYTDLTGTIQKISRDTGRTVQ 1017  
DB 1632 PNGSILESVIDTARVGEK-----RNGKDSFT-----FERDQNGQVTK 1670  
QY 1018 IKD--SSIKTTLYNDDLNRHIGSQVTDLATGHMLTTTVEFDGLNREIGRKLCDSSGHTLD 1075  
DB 1671 VKDLVNGVERTKTYDKADR-----VTS-AT-----DSRGGKID 1702  
QY 1076 IQQSWL-----KTQQLANRIVKLVGLQRTQYQSYDSRNLNQ--YKQDGAECPTDKY 1126  
DB 1703 ---WAVHDKANSYKELKEQTVTQGGY--TNKVSVD-YNTLDQNRIVTDGSO----- 1748  
QY 1127 GHSIVTQNTFYDVIYGNITACHTTFADGTEDHATKFE--ANP-----TDFCQTEVHT 1177  
DB 1749 ---TRFDYDDQGNV-----RTYTAGNGSGSTFNTDQANKIKDLVWGTSSNILLSERYE 1799  
QY 1178 HPDMPDIRLYKDKAG-----RVINITDNHGNTEFTYDYLGR-----LQ 1217  
DB 1800 YDQSGNRTKI KHEGAGGVKVTETNPYDPIQLLNEVL PNGTKSYTYDGFGRNRTSVKIE 1859  
QY 1218 NGQGS---VYGYDPLNRLVSQKTDTLCEL----- 1244  
DB 1860 NGKETSAIAATFNEGQNLVKFNGESLTYDVNGNRTSDGKYKTYWNEDDQIVAITKQGENN 1919  
QY 1245 ---YRETMVNEVRNGEMIRLLRTGETI----- 1274  
DB 1920 AFATYKDEDNRRTEKVVNGVQVTRYFYDGDGINSIPYETDNGTGVLRQVYVSADGARLANK 1979  
QY 1275 RASKVLLTGTDSSQOSVI--LTSQKQNLSEAYSAYGK-----HKSTANDASILGNGERA 1327



Db 1980 AQCQTLYHYNPRGDVVAMTNQDKVEVATYEDAWGNVLTSDTKGIAAD-NPFGYAGY 2038  
QY 1328 DPVSGVTHLNGYSDPTLMREHTPD-----SLSPFGAG-----GINPYSYCLGDPINR 1377  
Db 2039 DKEIGMYL-----IARYNPEHGVFLSVDPDPGDEDDPVTMNGYTYADNPNVM 2088  
QY 1378 SDPSGHLNQWATGIGMGIALLLTATGMAJAAAGGIAAIASTSTTALAFGALSVT 1437  
Db 2089 TDPDGKAW-----LVPVVIAGAMVAARFGAKYAI-----RYGAKYGGKAVKS 2131  
QY 1438 DITSIVSGALEDASPKASSILGWVSMGMAAGLAESAIGKGTKLATHLGAFAEDGENALL 1497  
Db 2132 -----GWDY-----GKKVAKSGWNKGSIAQKIPRIHKVGR---I 2163  
QY 1498 KSTSESRIRKMGVTRSLDRIVEIRNEBQVIKDHRSRGYTNFMKGQAILLVHGDGDFLY 1557  
Db 2164 KGDNKGGKGVGYISTTK-----KTGKRTYS-----SFEFHTPHNGHY 2203  
QY 1558 HTEGNKHNG-KGPYTR 1572  
Db 2204 HLQNKYSKYQKKNR 2219

RESULT 11

ID ADC01365 standard; protein; 1400 AA.  
XX  
AC ADC01365;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Enterohaemorrhagic E. coli 0157:H7-specific protein SEQ ID NO: 1410.  
XX  
KW enterohaemorrhagic; anti-bacterial.  
XX  
OS Escherichia coli; 0157:H7.  
XX  
PN JP2002355074-A.  
XX  
PD 10-DEC-2002.  
XX  
PF 24-JAN-2002; 2002JP-00015959.  
XX  
PR 24-JAN-2001; 2001JP-00112010.  
XX  
PA (UYTS-) UNIV TSUKUBA.  
XX  
DR WPI; 2003-451640/43.  
XX

PT Enterohaemorrhagic Escherichia coli 0157:H7-specific nucleic acid molecule  
PT and a polypeptide and its use, a polypeptide, a vector and a host cell.  
PS Claim 3; SEQ ID NO 1410; 2067pp; Japanese.

XX The invention relates to a novel enterohaemorrhagic Escherichia coli  
CC 0157:H7-specific nucleic acid molecule. A polynucleotide of the invention  
CC has anti-bacterial activity. The polypeptide can be used in detection  
CC and/or treatment of 0157:H7 infection. The nucleotide sequence of the  
CC genome of Enterohaemorrhagic E. coli 0157:H7 was determined. The present  
CC sequence represents an E. coli 0157:H7-specific polypeptide of the  
XX invention.

XX Sequence 1400 AA;

Query Match 4.1%; Score 365; DB 7; Length 1400;  
Best Local Similarity 20.9%; Pred. NO. 1e-15;  
Matches 284; Conservative 154; Mismatches 467; Indels 456; Gaps 65;  
QY 418 KTEW-RSADNSVRSPITETTFDESGNPLTKVI-----KOKKTKIISPSHWEY 466  
Db 218 RLTLYRREAAGDLAGEITGVTDGAGREFRLVLTQQAQRAEARKQHTASLSPDT---- 272

QY 467 YPPAGEVDCNCPPEPGYGTFRVKKIOTPPYDSEFKDDEKPIQYRYSLIGSQSHVTLKIE 526  
Db 273 -----PRPLSDSAFPDTPGTEY-----GPDGRIG----- 297  
QY 527 RHVSATQLLNSTLFOYNTDKSELGRLLKQTECTKGNGKTSYV-----VHKFTYTKOD 579  
Db 298 --LSAVLWTHDPAYPESLPGAPLARY-----TYTEAGELLAVYDRSNTQVRAFTYDAQH 349  
QY 580 DTLQOQSHSITTHDNFTIHRQSVRSRYTGRLFSDTDTKIVTQMSYDKGLRLLTRTLNSGT 639  
Db 350 PGRMVAH-----RYAGR-----PEMRVRYDDTGRVVVQLNPAGL 383  
QY 640 PYANTLTYDYELNLODDNRPPFVITTTD-VNGNQURNEFDGAGRHVSCQCKDSGDGKF 698  
Db 384 SY-----RYQYE-----ITVDSLNRREVLHTEGGAG--LKRVVKKELADG-- 425  
QY 699 YTIHTQYQDEQGRHHTSTYSYDLTNGR--QOQDPPDKVHLSMSKSYDNWQGIANTHWSYGV 756  
Db 426 -----SVTHSGYDAAGRLTAQTD-----AAGRRTYEGL 453  
QY 757 SEKITVDPITLTATKQLQSNNSNVQTKEVVITYTPSQPIQITLDFEAGHLQS-----CHT 812  
Db 454 N--VVSQDITDITPDGRETIFYNDGNQLTAVVSPDGLSRRAYDEPGRLVSETSRCDG 511  
QY 813 LTRDQGWDRVKE-----TDAIGCTIYQYDNNRVIQITLPDGTIVNRKYAPSTDTLIT 867  
Db 512 VIRYAYDNPHSELPAITTTDATGSTRQMTWSRYQLLAFTDCSGYQTRYEYDRFGQMTAVH 571  
QY 868 DIRVNGISLQQTFFGLSELTSQDG-GRVWATYYSAGNDQCPSTVITPDGQFIHVQYQP 926  
Db 572 --REEGIS--RYRRYDNRGLTSVKDAQGHETREYNAAGDL--TAVITPDGNRSETQY-- 634  
QY 927 ELDDA---VLQVASNEITQQFSYNPVTGALLKAVAGQSLTPIYYPGSLKMNENINDMKK 983  
Db 625 ---DANGKAVSTTQGLTRSEYD-LAGRITTLTNGSRSEPTYDA----- 667  
QY 984 MSYLTWLRGLENGYDGLTQIKISRDRTHGRVTQIKDSSIKTTLNVDLNRHIG----- 1037  
Db 668 LDRLVQQRGFD-----GRTQRYHYDLTGKLTQSEDEGLVTLWHYDESDRLTHRTVNGE 720  
QY 1038 -----SQVTDLATGHMLTTTVEFDGLAREIGRKLCDSSGHTLDI-----QQS 1079  
Db 721 PAEQWQYDEHGWLTSEIHLSEGHGVAVHYGDQKRLAGERQTVHNPETGELLWQHETEH 780  
QY 1080 WLKTOQLANRIVKLGVLQRTQYSYDS-----RNRL-----N 1112  
Db 781 AYNEQQLANRVTP--DSLPRVWLTGSGYLAGMKLGCTPLVEFTRDLRHRETVRSFGNN 838  
QY 1113 QYKCDGAECPDKYGH-----SIV-----TQNTFYDIYGNIT 1144  
Db 839 AVELTSTVTPA---GHLQSORLNSQVYDRDYDNDNDGLVRISGPRQTWEYGYSATGRLE 895  
QY 1145 ACHTTFACGTEDHATFKFANPTDPC--OLTEVHHTHPD-----MPDNIL-----KY 1189  
Db 896 SVRTLASD-----LDIRIPYATDPAGNRLPD--PELHPDSTLTAWPQN--RIADAHVYVRH 948  
QY 1190 DXAGRVINITD-----NHGNTENFYDTLGLRL-----QNGOGSV---YGYDPL-- 1229  
Db 949 DEYGLRTEKTDRIIPAGVIRTDDETHHHYDSQHLRVFYTRIQHGEPLVESRYLYDPLGR 1008  
QY 1230 -----NRLVSQKTDLTDCLEYLR----- 1247  
Db 1009 RMAKRVRRERDLTCWMSLSRKPEVTWYWGWDGRLTQTDTTTRIQTIVVPGSFTPLIRV 1068  
QY 1248 -----ETMLVNEVRNG-----EMIRLLRTGETIIAQRAS----- 1277  
Db 1069 ETENGEREKAQRRLAETLQOEGSENGHGVWPFAELVRLLDRLLEEIRADRVSSESAWML 1128  
QY 1278 -----KVLITGTDQO--QSVILTSQKONLSQBSAYSAYGKHKT 1313  
Db 1129 AQCGLTVEQLARQVEPYTPARKVHFYHCDHRGLPLALISEDONTAWRGDEYEWGQNLNE 1188  
QY 1314 ANDASI---LGYNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSY 1369

Db 1189 ENPYLHQPRLPQQHDESGLYNRRY--YDPLQGRVITQD---PIGLAGWNLNY 1243  
 QY 1370 CLGDPNRSPPSG-----HLSQAWTIGMGIAGLLT-----IATG-----GMAIAAGGI 1416  
 Db 1244 PL-NPIRMPLGLNLYQLLYDVHDDSYGTSIDITGSGDLISLGHAGLGVAFKKK 1302  
 QY 1417 AAIAASTSTALAFGALSVTSDITSIVSGALEDASPKA--SSILGWVSMGMGAALAES 1474  
 Db 1303 GEMLSDICIIYATAGHAGIGGGINAITYYSKSLTSGVSNVSG-VTVGGVGG----- 1356  
 QY 1475 IKGTKLATHLG-AFAEDGKNALLKSTBSRSRIKWGVTRSL 1514  
 Db 1357 -----HFAVTVYVDVNP--ESSTESVGIGAGVDAV 1386

## RESULT 12

ADG30698  
 ID ADG30698 standard; protein; 1510 AA.

XX AC ADG30698;

DT 26-FEB-2004 (first entry)

XX Xanthomonas axonopodis pv citri plant pathology-related XAC3245 protein.

XX Xanthomonas microorganism; plant; pathology; bacterial pest; Xac; Xcc;

XX XAC.

XX Xanthomonas axonopodis pv. citri.

XX WO2003089647-A1.

XX 30-OCT-2003.

XX 22-APR-2003; 2003WO-BR000060.

XX 22-APR-2002; 2002US-0374620P.

XX (AMPA-) FUNDACAO AMPARO A PESQUISA DO ESTADO.

XX Da Silva ACR, Farah SC, Quaggio RB, Reinach FDC, Ferro JA;

XX De Oliveira JCF, De Laia ML, Setubal JC, Furian LR;

XX WPI; 2003-865444/80.

XX N-PSDB; ADG30697.

XX New nucleic acid molecule from a Xanthomonas microorganism, useful in

XX determining the presence of Xanthomonas bacteria in a sample.

XX Claim 8; SEQ ID NO 55; 145pp; English.

XX The invention relates to a novel isolated nucleic acid molecule from a

XX Xanthomonas microorganism where the nucleic acid molecule is associated

XX with pathogenicity caused by the Xanthomonas microorganism, or its

XX variant, that causes reduced or enhanced pathogenicity. The nucleic acid

XX of the invention may be useful in detecting the presence of Xanthomonas

XX bacteria in a sample, as well as in plant pathology, for example, for

XX identifying nucleic acid molecules and proteins involved in pathology

XX caused by bacterial pests. The current sequence is that of the

XX Xanthomonas axonopodis pv. citri (Xac) plant pathology-related XAC

XX protein of the invention.

XX Sequence 1510 AA;

XX Query Match 4.1%; Score 361; DB 7; Length 1510;

XX Best Local Similarity 21.2%; Pred.No. 2.2e-15;

XX Matches 304; Conservative 158; Mismatches 466; Indels 504; Gaps 75;

XX 144 VLTGFNNNAFDLK-----VPKKLNPAGHAIYIDWNFEATQRLNRIYDLDG 191

XX 250 ILTRPNGNFYTFKKSANGTWTPDWDVRETLSEVDNGTLTGQVATDSDSKEQF--DLDG 307

QY 192 HDIPLLNLEYQGLIKTILTLFPQGEKYRTELFLNRLNLSIHN-----F 236  
 Db 308 ---KUTGISY-----TDGQHLTLYAGROLQSVTDTRGRLLLFAYOADRIT 350  
 QY 237 SSGNENPLTWSRGYPIGKNGILGOWITSMTPAGGLKETV---NYSNNQGHFFQS--- 290  
 Db 351 QVGLPDGMVLAVGYDSQARL---RSVLTQTAAGAVASAIAGYDGN---ARFFDALTT 402  
 QY 291 -----ANLPLVPLVTLMKQVPGAGOPAI-QAEYSY-----TSHNYVGG---GSN 330  
 Db 403 HRDEQDQVAVTYDAQQRVVRSVHGDPTGKIDEATIAYSNGNTSVSNALGNVAVTRTGIS 462  
 QY 331 GIWNNKLNLYGLMTEYNYGSTESR-----YKQEGHDOIVRIERTYNNVHLLTSECKQ 385  
 Db 463 KLGQAKVTAVQGLCECTVGFAPKRSYDNGYVPOEVDVFGVATDWRVNNRGLLASKIEA 522  
 QY 386 QN---GYIQTETAYYAIIGHNFDSPQSQFOLPKTKTETWRSAD-----NSYRSE 432  
 Db 523 DNSSGGQKRTLQT-----DWHPS-FRVP-TDQRTYDASDILVARTSWTYSNRQQA 570  
 QY 433 ITETTFDESGNPLTKVIKDKTKIISPSHWEYYPAGEVNDNCPPEPYGTRFVKKI 492  
 Db 571 LTVSRTPSGGP-TRV---TTQRYCEDSD---TAAG---NC----- 601  
 QY 493 TPYDSEFKDDPEKFIQYRYSILGSSQSHVTLKIEERHYSATOLLNLTLPQYNTDKSELGRL 552  
 Db 602 ---KQTECT-----KGNGKTYSVVHKFT-YTKQDD-----TLQQSHSITTHDNFTIH 597  
 Db 628 ADDASCTASASTCPRKGRDRWKVTNALQQTTEYLAYDAGRPLSIKIDNGIVT--DYTYH 685  
 QY 598 -----RSQVRSRTGRLFSDTDKDIVTQMSYDKLGRLLTRTLNSGTPYANTITYDEL 651  
 Db 686 PRGMLTASKVRG-----ADASSEADRIIDYWPTG--LVRQVTPQDGAFTAYTD-AA 737  
 QY 652 NNLODNRPPFVITTTDVNGNLRNEFDGGRHYSQCLKDSGDGKFYTHQOQYDEGR 711  
 Db 738 HRLTD-----ITDNAGTVHYTLDNAGRVKEDTKDAAATLK--RTLVRVYNQLGQ 786  
 QY 712 HHT--STYSD-----YLTNGRQQTDPDKVHLSMSKSYDNNMGOI-----ANTHWS 753  
 Db 787 LKTQATAASDPTDFAYDANGNATKVTDALATATQSEYDPLNRLSHTLQDVAGIKADTKFA 846  
 QY 754 YGVSEKIT--VDPILTATKQLOSNVQ-----TGKEVTTYTPSQOPIQITLFDKACH 806  
 Db 847 YDALDNLTKVTDPKGLDITYDNGFGBELVKLTSPDTG--VTSYT-----YDSAGN 894  
 QY 807 LQSCHTLTRDGRVRRKETDAIGQCTIYQVDNVRVITLPLDGTIVNRKYAPFSTDTLI 866  
 Db 895 RAT-----QTDARGNTTAYSDALNRLTKVYPTS-----SLDVTY 930  
 QY 867 T-DIRVNGISLGQQTDFGLSRLTQSDGGRVWYATYSAGNDQCPSTVITPDQFIHYQY 925  
 Db 931 TYDVQTACTSG-ETF-SIGRLTKMQDGGAITQCYNRFGD-LVRKYQTSNGTALVLRD 987  
 QY 926 PEL-----DDAVLQVASNE--ITQOFSYNPVTGALLKVAEQSLTPIYYPGRL 973  
 Db 988 YTVGGQLRRMTYPDGAVVDVYVRNAQQTQVGVTPAGGS--RQVLLGNA---TYPPGPA 1042  
 QY 974 KMENINDMKMSYLVLT-----RGLNGYTDLAGTIQIKISRDTHGRVTQIKDS 1021  
 Db 1043 AGWTYNGNRTLARQYDLDRPQAIQDTRPGGLDVGF-----GFDPAIGNLTALTPA 1092  
 QY 1022 SIKTT---LNYDDLNRHIGSQVTDLATGHMLTTVPFDGLNREIGRKLCDSSGHTLDIQ 1078  
 Db 1093 GNTTPEIGLGYDALGRLTG--LTDGVTGTVI-----DG----- 1123  
 QY 1079 SWLKTQQLANRIVKLVNGLQRTQOYSYDSR-NRLNQYKCDGAECPDCKYHSIVTONFTY 1137  
 Db 1124 -----YSYDATGNRL-----STKVG--TATQIYTY 1146  
 QY 1138 DIVGNITACHTTADGTEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYKAGRVIN 1197



QY 1170 Q-----LTVHHHPDMPDNIRLKYDKAGRVINITDNGHTENFTYDTL 1213  
Db 578 QGETAQAGAGSVVFFNRIITSYRGLH-----YRYDEYGRVVEKGRNG-IGHYRWDAE 628  
QY 1214 GRL-----QNGGGSVYG--YDPLNRLVSKQDTLCELYRYETMLVNEVRNGEMIRLLR 1265  
Db 629 HRTEVAVTRGGTVRRYGVYDAPGRVRVEKHELDAGKFPYRNTTFIWDGMRLAQECLGR 688  
QY 1266 TGETIIAQOR-----ASKVLLTGTD-SQGSVLTSDKONLSQEA-YSAY 1307  
Db 689 SSSLYISDRGSHEPLARVDRAAPGEADEVLYHTDVGAPBMTDGGNIWEAGYQW 748  
QY 1308 GK---HKSTANDASILNGERADPVSGVTHLNGVRSYDPTLMRPHTPDLSLSPFG-AGG 1363  
Db 749 GNLTHEKETRPVQONLRFQGYLDRETGLHY--NLVRFYDPDYGKFGSD---PIGLAGG 803  
QY 1364 INPYSYCLGDPINRSPSG 1382  
Db 804 INDYQYA-ENPLSYIDPLG 821

RESULT 14

ABU19676  
ID ABU19676 standard; protein; 1515 AA.  
XX AC ABU19676;  
XX 19-JUN-2003 (first entry)  
DE Protein encoded by Prokaryotic essential gene #203.  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX Borrelia cepacia.  
XX WO200277183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948997.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX N-PSDB; ACA23546.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 25; SEQ ID NO 47600; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 1515 AA;

Query Match 3.9%; Score 350; DB 6; Length 1515;  
Best Local Similarity 22.1%; Pred. NO. 1.3e-14;  
Matches 232; Conservative 126; Mismatches 314; Indels 380; Gaps 59;  
QY 481 YGTRFVKKIIQTPYDSEFKDDPEKFIQYRYSRLIGSQSHVTLKIEERHSATOLLNSTLF 540  
Db 543 FQGGFYVAYDEHGMWMTQWRDQTDVRYRYDTAG-----RVVE-----TGRQ 586  
QY 541 QYNTDKSELGLKQTECTGKENGKTVSVHKFTYTKODDTLQOSSH-----ITHDNFTI 596  
Db 587 GYHT-----GRFIVEAGCT-----RVIDVGEWYAYNDEGLVTAETDPLGHCYSEWEL 636  
QY 597 HRSQVR-----SRY--TGRIFS-----DTDKDIVT-----QMSY 624  
Db 637 GRLMARIDPLGRTRDYDERGQLTSVVSSGRTVDFDYDDEQRLTGARLPNGGKIKLEY 696  
QY 625 DKGLRLRLTNSGTPYANTLTIDY-----EL-----NNLODDNRPFFVITTDV-- 669  
Db 697 DHLRLRIART-----EPDGNKTTYRGPFGELRLRVVQGRRETRLDYDOR-----LRUTDIEL 748  
QY 670 -NGNQLRNEFDGAGRHSQCCLKSDGDKFYTHITQYDEQGRHHTSTYSYDLNAGR--- 725  
Db 749 PTGARFRKKIDALGR---LLEETSPDG-----HVTRYD-----YADGPANPGLL 790  
QY 726 -QOTDPPDKVHLSMSKSYDNWQJANTHWSYGVSEKITV---DPITLTKOLOQSNVNVQ 781  
Db 791 SAVTRPD-----GSVSRARYN---SESLPVEWIDPLGRTIQR----- 824  
QY 782 TGKEVTVTPSQQPIQITLPDEAGH-----LOSCHTLTRDGDWRVRK 823  
Db 825 -----TYGPPD--LITASIDAAGHATRFYDHAHTRLTKVINALGETTYTYRYDAAGRLAA 876  
QY 824 ETDAIGQCTIYYQYDYNRVIIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLQ--QTF 881  
Db 877 EIDWGRATEYDRDAVGRLLTKTLPDG-----QWRVYTY 910  
QY 882 DLSRLTQSQDGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASENIT 941  
Db 911 DASRLLEIDAGDVKYLARYDASG-----RLASAEVQGEHTHTV 949  
QY 942 QQFSYNPVTGALLKAAVAGQSILTYIYPSGRKLKMNINMKMSYLTWLRGLENGYTDLT 1001  
Db 950 -RFAYDR-NGRLIGEDQHGELLRHVYDADGQRRU--RMTPRRETTYA----- 993  
QY 1002 GTIQKISRDRTHGRVTOIKDSSIKTTLNLYDNLNRHIGSQVTDL-----ATG-----HMLT 1050  
Db 994 -----DVSALTQVG---QLTIRRDGLGREIGREAGDFVAQQYQYDALGRIRRRQIAG 1041  
QY 1051 TTVEFDGNEIRGKRLCDSSGHTLDIOQSWLKTQOLANRIVKLVGLVQRTQEQYSYDNRN 1110  
Db 1042 PAVAFDALQADPARAL-----EQLTRQV-----YHYDAAGQ 1072



```

Db 715 AERWQYDE-RGWLTDISHISEGHRVTVHYGDSKGRLAGSEHLTVHHPTQNELLMQHETRH 773
QY 987 LWTLRGLNGYTDLTGTIQKISRDTGRVTQIKDSSIKTTLNVDLNRHIGSQVTDLATG 1046
Db 774 AYNAQGLAN--RCIPDSLPAVEWLTYG-----SGWLSGMKIG 808
QY 1047 HMLTTTVEF--DGLNREIGRKLCDSSGHTLIDIOOSWLKTQQLANRIVKLVGLVQTEQYS 1104
Db 809 D--TPLVEYTRDLRHRETLR-----SFGRYELTAYTPAGQLOSO--HLNSLLS--DRDYT 858
QY 1105 YDSRNLNQKCDGASCPTDKYCHSIVTONFTYDIYGNITACHTTPAD-----GTEDHA 1158
Db 859 WNDNGELIR-----ISSPRQ-----TRSYSYTTGRLTGVHTTAANLDIRIPYTTDPA 906
QY 1159 TPKFANPTDPCQLTEVHHTHPD-----MPDN-----IRLKYDKAGRUNITD----- 1200
Db 907 GNRLPDP-----ELHPSALSMWPDNRIDARDAHYLYRYDRHGRUTEKTDLIPEGV 956
QY 1201 ---NHGNTENFTYDTLGRQLNGQGSVYG-----YDPLNRLVSQKTDTLDCEL----- 1244
Db 957 IRTDDERTHRVHYDSQHLVHYTRTQYAEPLVESRYLYDPLGRRVAKRVWRERDLTGWM 1016
QY 1245 -----YY-----RETMVNE-----VRNGEMIRLLR-----TGETIIAQORA--- 1276
Db 1017 SLSRKPQVTWYGWDGDRLTIONDRTRIOTIYQPGSFTPLIRVETATGELAKTORSLAD 1076
QY 1277 -----SKVLLTGTDSQSQSVILTSQKQLSOEA----- 1303
Db 1077 TLQSGGEDGGSVVFPFVLQMLDRLESEILA---DRVSESRRLWASCGLTVAQWQSQM 1133
QY 1304 -----YSAYGKHKSTANDAS-----ILGYNG 1324
Db 1134 DPVYTPARKIHLYCHDRGLPLALISTEGTTAWYAEYDEWGNLLNEENPHQLQQLIRLPG 1193
QY 1325 ERADPVSGVTHLNGVRSYDPTLMRHTPDSLSPPG-AGGINPYSYCLGDPINRSDPSG 1382
Db 1194 QQYDEESGLYY--NRHRYDPLQGRYITOD---PIGLKGMNFYQVPL-NPISNIDPLG 1246

```

Search completed: February 9, 2005, 18:17:39  
Job time : 224 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2005, 18:10:24 ; Search time 50 Seconds

(without alignments)  
2497.758 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879

Sequence: 1 VIKELKLPFRITMSDNEF.....PRKILGRTEKTVKPTFRP 1673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/prodata/1/iaa/5A COMB.pep:\*
- 2: /cgn2\_6/prodata/1/iaa/5B COMB.pep:\*
- 3: /cgn2\_6/prodata/1/iaa/6A COMB.pep:\*
- 4: /cgn2\_6/prodata/1/iaa/6B COMB.pep:\*
- 5: /cgn2\_6/prodata/1/iaa/PCTUS COMB.pep:\*
- 6: /cgn2\_6/prodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1638	18.4	1584	3	US-09-251-645-6
2	339	3.8	1426	4	US-09-492-709A-340
3	331.5	3.7	1377	4	US-09-711-164-467
4	317	3.6	998	4	US-09-252-991A-28424
5	313	3.5	1627	4	US-09-328-352-6604
6	306	3.4	1586	4	US-09-543-681A-5329
7	294	3.3	1439	4	US-09-543-681A-7560
8	292	3.3	1626	4	US-09-252-991A-23805
9	285.5	3.2	1596	4	US-09-328-352-5542
10	277.5	3.1	974	4	US-09-252-991A-23640
11	271.5	3.1	1665	4	US-08-851-567B-61
12	238.5	2.7	1043	4	US-08-851-567B-61
13	235.5	2.7	804	4	US-09-328-352-5545
14	223	2.5	1128	4	US-09-252-991A-31032
15	222.5	2.5	3290	4	US-09-328-352-5486
16	220.5	2.5	1183	3	US-08-447-031A-2
17	218.5	2.5	2123	3	US-08-968-685A-10
18	214	2.4	2504	4	US-09-328-352-5821
19	212.5	2.4	954	3	US-09-251-645-12
20	211.5	2.4	10182	3	US-09-134-001C-3159
21	207	2.3	2314	3	US-09-268-347-49
22	202.5	2.3	2777	4	US-09-543-681A-6124
23	201	2.3	1739	4	US-09-540-236-3739
24	201	2.3	2057	4	US-09-499-203-2
25	199.5	2.2	1092	4	US-09-147-405B-15
26	198.5	2.2	1004	3	US-09-268-347-30
27	197.5	2.2	1726	4	US-09-700-227-2

28	194.5	2.2	5024	4	US-09-710-279-2964	Sequence 2964, Ap
29	190	2.1	2385	4	US-09-543-681A-6304	Sequence 6304, Ap
30	188.5	2.1	2736	4	US-09-252-991A-30227	Sequence 30227, A
31	185.5	2.1	1166	4	US-09-200-650B-7	Sequence 7, Appli
32	185.5	2.1	2383	4	US-09-492-709A-302	Sequence 302, App
33	184.5	2.1	2142	4	US-09-540-236-3459	Sequence 3459, Ap
34	184	2.1	3696	3	US-09-134-001C-5080	Sequence 5080, Ap
35	183	2.1	1222	4	US-09-206-942-37	Sequence 37, Appl
36	183	2.1	1228	4	US-09-206-942-34	Sequence 34, Appl
37	182.5	2.1	1565	4	US-08-851-567B-59	Sequence 59, Appl
38	182	2.0	1861	2	US-08-790-912-4	Sequence 4, Appli
39	181.5	2.0	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
40	181	2.0	1228	4	US-09-463-402-2	Sequence 2, Appli
41	181	2.0	1228	4	US-09-889-572-2	Sequence 2, Appli
42	181	2.0	1228	4	US-09-117-447-2	Sequence 2, Appli
43	180.5	2.0	930	3	US-09-134-001C-5314	Sequence 5314, Ap
44	180.5	2.0	930	4	US-09-386-962C-10	Sequence 10, Appl
45	180	2.0	1683	3	US-08-755-587-183	Sequence 183, App

## ALIGNMENTS

### RESULT 1

US-09-251-645-6  
; Sequence 6, Application US/09251645  
; Patent No. 6281413  
; GENERAL INFORMATION:  
; APPLICANT: Kramer, Vance C.  
; APPLICANT: Morgan, Michael K.  
; APPLICANT: Anderson, Arne R.  
; APPLICANT: Hart, Hope  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Dunn, Martha  
; APPLICANT: Chen, Jeng S.  
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS  
; FILE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS  
; CURRENT APPLICATION NUMBER: US/09/251,645  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 1584  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-251-645-6

Query Match	18.4%;	Score 1638;	DB 3;	Length 1584;
Best Local Similarity	31.0%;	Pred. No. 1.3e-117;		
Matches 493;	Conservative 233;	Mismatches 682;	Indels 182;	Gaps 54;
QY	19	EPTQANNPTSAVSGVDPRTGTYNIQTILGHVGN--GNLQFTLPLTISYSPKNTDYG	76	
Db	5	DIYSNAFPGSYINTGVDPRTGQYSANINIIILRPNNVGNSEQT--LSLSFSLPTLTNNG	62	
QY	77	FGIGNFGLSVYDRKNSLLSSTGENYK--VIEDTKVKLOQKLDNIRFERKDLKENCY	133	
Db	63	FGIGWRFSLTLDIKTLTFRSANGQPKCKPLPPNNNDLSFKDKLKDLRVYK-LDSNTF	121	
QY	134	RIIHKSGDIEVLTGFNNNAFDLKVPKLLNPAGHAIYIDMNFETQPLRNRIYDDLDGHD	193	
Db	122	VYVKNKGIIEILKRIGSS--DIAKTVALEFPDGEAFDLYNSR-----	162	
QY	194	IFLLNLEVOGLTKITLT--FPQCKEGYRTELRL--NRQLNSIHNSLGNENPLTWSFGYT	251	
Db	163	FALSEIKYRVTKTYTKLYNSGNN---CYSVEYPPDNNISAKIAFDYRNDYLTITVTPYD	219	
QY	252	PIGK-----NGILQWITSMTPAGGLKETVNSYNNQGHFFPQSANLPVLVY	298	
Db	220	ASGPIDSRPFKMTYQTLKGVF-PVISTFTPTGYVELSVYKEN--GH---KVTDTYIY	273	
QY	299	VTLMKQVFGAGQPAIQAEYSYTS-HNYVGGSGNGI-WNNKLDNLYGLMTIYNGSTESRR	356	



Db 274 AALATIOPGCGPAVSKSYEYSSVHFLGYSRGTSFDSQDNLYLVGKYYSIE--- 330  
QY 357 YDKKEGHDQIVRIERTYNNHLLTSECKQONGYIOTTETAYVAIIGHNFDSPQFQIP- 415  
Db 331 -RVLGQSVVSVIERVFNKFLMTKEAKTDQNKRIITTEITYNEDLSKSPSEPNLQOPS 389  
QY 416 KTKTEWRSADNSYSEIETETTFDSGNPLTKVVKDKTKLIISPTHWVYPPAGEVDN 475  
Db 390 RVLTRYTDIOTNSRETNVNIKSDWGNLIL-ITETSGIQK-----EYVYPPVNGEKS 442  
QY 476 CPPEPYGFRFVKKIIQTPYDSEFKDDPEKFIQYRY-SSLIQSO-----SHVTLKIEE 526  
Db 443 CPADPLGFSRFLKSVTKQSGPDAQSVANKVIHYIYQKPPFTTGAVVKEYYSKVSETIDN 502  
QY 527 RHYSATOLLNLSLTFQYNTDKSELGRLLKOTECTKGENGKTSYVVKHFTYTKODDTLQOSH 586  
Db 503 KIARTFSYVNSP-----TSKSH-GSLAKITSVMNQ-----QTVVTFKYEYSESEMTTNA 551  
QY 587 SITTHDNFTIHRSQVRSRYTGLFSDTDKDIQTQMSYDKLGRLLTRTLNSGTPYANTIL 646  
Db 552 YTVGFDGAMESKNVTSIYTHQLKVDVNVHVTQSDYLLGRITGQIIDPGTAREIKRN 611  
QY 647 YDYELNNLQDDNRPPFVITTTDVNGNQLRNEEDGAGRHVSQCLKDSG-----DGKF 698  
Db 612 YVYQYPGDENDFWP-VMEVDSQGVRRKTHVDGMRICISIEQDDGAWGTSGIYQGTY 670  
QY 699 YTHIQOQDEQRHHTSTYSYLTNGROQTD-----DKVHLSMSKSYDNWQJANTHWS 753  
Db 671 RKVLARQYDVLGQLSKEISNDLWN--LSANPLVRLATPLVTKYKYDGMGNLYSTEYS 728  
QY 754 YGVSEKIVDPLTLATK-----QLOSNNSNVQTKVTTTPSQOPIQIILFDEAGHL 807  
Db 729 DGRILEIHDPIRTITQGVKGLMNLQOQNF-----EQPASIKAVYPDGTI 776  
QY 808 QSCHTLRDGDWRVRKETDAIGCCTIYQYDNTNRYIQTLPDGTIVNRKYAPFSTDLIT 867  
Db 777 YSTRYVYDGFRTVETDAEHAQIGYDVFDRIVKVTLPDGTILESAFSSHEELIS 836  
QY 868 DIRVNGISLQOQTFGLSRLTQSGGRVWANTYSAGNDQCPSTVITPDQPIHYQOPE 927  
Db 837 ALNVANGTQGLAVYDGLGRVISDTVGRKTEYLYGPGDK-PIQISITPSHNKNQMDLYY 895  
QY 928 LODAVLOVANSIEITOQFSNPNVTGALLKAVAGQSITPI-YVPSGRLKXEN-INDMKWS 985  
Db 896 LGSVMSKFTTGTDQONFRHSHKIGLLSA-SEGVQTNVSPPSGVLORESFLRNKPFIS 954  
QY 986 ---YLWTLRGLNGYTDLTGTIQTQKISRDRTHGRVTOIKDSSIKTTLNLYDDLNRHISQVTD 1042  
Db 955 SGEYLVTWGLIQRHKDSFGHNHVSYYDAQGRLVKTEQDAQYATFEYDNGVGLITTTKD 1014  
QY 1043 LATGHMLTTVTFDGLNREIRGRKLCDSGHTLIDIOQSWLKTQOLANRIVKLVGLQRTQ 1102  
Db 1015 TTSLSQVTKIEYDAFDREIKRSLISDFSIOV-ITLSYTKNNQISQRTSISDGVVMKNER 1073  
QY 1103 YSYDSNRNLNOYKCGAECPTKYCHSIVTQNFYDIYGNITACHTTADGTEDHATKFE 1162  
Db 1074 VOYDNNQRUSQYQCEGSESPIDHTGRVLNQIYHYDQWGNIKRLNLTTRDGKET-VDYHP 1132  
QY 1163 ANFTDPCQTEVHHTHPDPMDNIRLYDKAGRVINIT-DNHGNTENFYDTLGRIL---QN 1218  
Db 1133 SQ-ADPTQLIRI-----TSKQOQIELSYDANG---NLTRDEKQOT--LIYDQNNRLVOVKD 1182  
QY 1219 GOGSV---GYDPLNRLVSQ---KDTYLDCELYYRETMLVNEVRNEMIRLLRTGTETIAQ 1273  
Db 1183 RLGNLVCSYQYDALNKLTAQVLANGTVNRQ-HYASGKVTNIOQLGDEAITLSSDKORICH 1241  
QY 1274 OPA-----SKVLLTGTDSQOSVILTSDKQNLQSOEAYSAYCKHKSSTANDASILGNGERADP 1329  
Db 1242 QSAKNGQSVYQYGHIDHNSVTJASQENELMALSVTPYGFRLI---SSLFELNQAQVDP 1298  
QY 1330 VSGVTHLGNVRSYDPTLMRFHTPDSLPFGAGGINPYSYCLGDGDPINRSDPSGHLISQAW 1389

Db 1299 VTGWYFLNGYRVFNFVLMRFHSPDSWSPFGRGGINPYTYCQGDPINRIDLNGHLSAGGI 1358  
QY 1390 TGIWGIAGILLTAT--GMAIA-----AAGIAAAIASTSTTALAFGALSVTISITSIV 1443  
Db 1359 LGIVLGAITGIIVISLGAAGAAISAGLIAAGALGAIASATSALAVTATVIGLAADSIGIA 1418  
QY 1444 SGALEDASPKASSILGWSMGMAAGLAESAIGKGTKLATHLGAFAEDGENALLKST--SE 1502  
Db 1419 SAALSEKDKPKTSGLINWISAGLVUSFGISAI-----TFT-----SSLVKSARS 1463  
QY 1503 SSRKMGVTRSLDRREIVRNEBQVIKOHSR 1532  
Db 1464 SOAVSAGVIGSVPLEF-----GEVASRSSR 1488

## RESULT 2

US-09-492-709A-340  
; Sequence 340, Application US/09492709A  
; Patent No. 6720139  
; GENERAL INFORMATION:  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Trawick, John  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Froelich, Jamie M.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
; FILE REFERENCE: ELITRA.001A  
; CURRENT APPLICATION NUMBER: US/09/492,709A  
; CURRENT FILING DATE: 2000-01-27  
; NUMBER OF SEQ ID NOS: 485  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 340  
; LENGTH: 1426  
; TYPE: PRT  
; ORGANISM: E. Coli  
US-09-492-709A-340

Query Match 3.8%; Score 339; DB 4; Length 1426;

Best Local Similarity 20.2%; Pred. No. 1.1e-16;

Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;

QY 572 KPTTKQDDTLQQSHSITHDNFTIHRSQVRSRYTGLFSDTDKDIQVQMSYDKLGRLL 631

Db 317 RYTYTEAGELL---AVYDRSNTQVRAPTYDAQHPGRVMAHRYAGRPPEMYRYDDTGRVV 372

QY 632 TRLNSGTPYANTLYDYELANNLQDDNRPPFVITTD----- 668

Db 373 EOLNPAG-----LSYRY---LYEQDR---ITVDSLNRREVLTHTGGAGLKRVRKKE 418

QY 669 -VNGNQLRNEPDGAGRHSQCCLKDSGDGKGYTIHTQOYD-----EGCRHHTSTYSYDL 721

Db 419 LADGSVTRSGYDAAGRLTAQ--TDAAGRTYGLNVVSGDITDITTPGRETIFYND-- 474

QY 722 TNGRQOT---DPDKVHLSMSKSYDNWQJANTHWSYGVSEKIVTDPIITLTKQLQSNRN 778

Db 475 --GNQLTAVVSFD--GLESREYDEPGRLV-----SE 502

QY 779 NVQTKGYT-TYTPSQOPIQITLDEAGHLSCHTLTREDGWRVRKETDAIGCCTIYQYD 837

Db 503 TSSGSETVRYRYDDAHSELPAITTTDATG---STROMTWSRYGQLLAFTDCSQYQTRYBYD 559

QY 838 NNNRVIQITLPGDTIVNRKYAPFSTDTLITDIRVNGISLQOQTFDGLSRLTQSDG-GRV 896

Db 560 RFGQM-----TAVHRE-----EGISL-YRYDNRGLTSVKDAQGRE 595

QY 897 WAYTYSAGNDQCPSTVITPDGQFIHYQYQ----- 925

Db 596 TRYSEYNAAGDL--TAVITPDGNNRSETQYDAMGKAVSTTQGGLTRSMEYDAAGRVISLTNE 653

```
QY 926 -----PELDAVLQASNEITQOFSYNNPVTGALLKAVAGQSLTIPIYPSGRKME 976
Db 654 NGSHSVFSDALDRLLVQOQGGFDGRTQRYHD-LTGKLTQSEDEGLVILMYIDESDRTHR 712
QY 977 NINDMKMSYMLTWLRLGLENGYDLTGTIQKISRDTGRTVQIKDSS-----IKTLLNYDDL 1032
Db 713 TVNGEPAEQWQY-----DGHGWLTDISHLSEGRHVAHYGYDDK 751
QY 1033 NRHIG--SQVTDLATGML-----1049
Db 752 GRLTGBCQTVENPETGELLWQHETHKAYNEQGLANRVTPDSLPPVEWLTGSGYLAGMKL 811
QY 1050 -TTTVEF--DGLNREIGRLCDSSGH--TLDIQOSWLKTOQLANR-----1089
Db 812 GGTPLVEYTRDLRHRTVRSFGSMAGSNAAYELTSTYTPAGQLQSHLSLVYDRDYGWS 871
QY 1090 ----IVKLGVLQRTQOYSVDSNRNLNQVKDCAEC-----PTDKYGHSI-----1130
Db 872 DNGDLVRIISGPRQ--TREYGYSATGRLESVRTLAPDLDIRIPYATDPAGNRPDPPELHPS 930
QY 1131 -----VTQN-----FTYDIYGNITACHTTFPADG--TEDHATFKFANPTDPCQLTEV 1174
Db 931 TLTWFPDNRIAEADAHVYVRHDEYGRLTETKTRIPAGVIRTDDBT-----975
QY 1175 HHTHPMPDNIRLKYKAGRVINITD-NHGN-----TENFTYDTLGR-----1215
Db 976 HHYH-----YDSOHLVFTYRIQHGEPVLESRYLYPLGRMAKRVWRERDLTG 1025
QY 1216 ----LQNGQSVGVYDPLNRLVSKQTDLTDLCELYR-----1247
Db 1026 WMSLSRKEVTVYGMWGQ--DRLTIVQTDTRITQTVPEPGFTPLIRVETENGEREKAQRS 1084
QY 1248 --ETMLVNEVRNG-----EMIRLL-RTGETIIA-----QQRASKV 1279
Db 1085 LAETLQOEGSENGHGVVFPABLVRLLDRLEERADRVSSESRAWLAQCGLTVEQLARQV 1144
QY 1280 LITGDSQOS-----VILTSKONLSOEA--YSAYGKHKSSTANDASILGN-----1323
Db 1145 EPEYTPARKAHLVCHDRHGLPLALISEDGNATWSAEYDEWGNQLENPHV--YQPYRL 1202
QY 1324 -GERADPVSGVTHLNGYSYDPTLMEHTPDSLPPG--AGGINPYSYCLGPIINRSDPS 1381
Db 1203 PGQOHDEESGLY--NRHRYDPLQGRYITQD---PMGLKGGWNLQYPL-NPLOQIDPM 1256
QY 1382 GHLSWAQWTGIGMGIAGLLLTATGMAIAAAGIAAIASTSTTALAFALSVTSIDITS 1441
Db 1257 GLL--QTWDDARS-----ACTGGV---CGVLSRIIGPSKFDSTADALD-----1296
QY 1442 IVSGALEDASPKASSILGVWGMGAGLAESAIGKGTKLATHLGAPASDGENALLKSTS 1501
Db 1297 ----ALKETQNRS-----LCNDMEYSIGIVCKDTNG-----KYFASKAETDNLK--K 1336
QY 1502 ESSRIKWGVTSLDRIVNEEQVLIKDHSGYTDNFMKGBCQAAILVHGDK--DGFLYHT 1559
Db 1337 ESYPLKRCPTGTRVAAVHTHG---ADSHGDYVDEFFSSSDKNLVRSDNNLEAFYLAT 1393
QY 1560 EGNKH---NGKGPY 1570
Db 1394 PDGRFEALNNKGEY 1407
```

## RESULT 3

```
US-09-711-164-467
; Sequence 467, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
```

```
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 467
; LENGTH: 1377
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-711-164-467

Query Match          3.7%; Score 331.5; DB 4; Length 1377;
Best Local Similarity 20.8%; Pred. No. 4e-16;
Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;

QY 211 LFPQOKEGY-RTELPLNRQLNSIHNFSLGNENPLTWSEFYTPIGKNIGLQWITSMTAP 269
Db 126 LFPGE-DGYSRSESLWLR--GGVAKLDGHRLAALWQ-----160
QY 270 GGLKETVNY-----NNQG-----HHFPQSANL---PVLPIYVILMKQVPGAGQ- 310
Db 161 -ALPEELRLSPHYLATNSPOGPMWLLGCERVPEADEVLPAPLPPYRLVTLGLVDRFGRT 219
QY 311 PAIQAEYSYTSNYYVGGSGNGIWNKLDNLYGLMTEYNYGSTESRYKDEGHQDQIVRIE 370
Db 220 QTFHRAAGEFGEITGVTDGAWRH-----FRLVL-----TTQAQR-----255
QY 371 RTNNYHLILTSCKQNGYIQTETAYYAIIGHNFDSPSQFOLPKT---KTETWRSADN 427
Db 256 -----AEARQQ-----AISG---GTEPSAF---PTLPGYTEYGR--DN 287
QY 428 SVRSEITETTFDESNGPLTKVIKDKTKIISPSHWEYYPAGEVNDNCPPEP---YGFT 484
Db 288 GIRLSAVMLTHDPE-----YP-----ENLPAAPLVRYGWT 317
QY 485 -----RPFVKIITPYDSEPK-----DDPEKFIQYRYSLSIGSQSHVTLKIE 525
Db 318 PRGELAVVYDRSGKQVRSFTYDDKYRGNRVAHRHTGRPE--IRYRD---SDGRVTEQLN 372
QY 526 ERHSATQLLNSTLFOYNTDKSELGRLLKQTEC--TKGENGKTYSYVHVHFTYTKQDDTLQ 583
Db 373 PAGLSVT-----YQEKDRITITDSDLRREVLTHTQGEAG-LKRVVK-----413
QY 584 QSHSITTHDNFTIHSQVRSRYTGLRFSDDTKDIVTQMSYDKLGRLLTRTLNSGTPYAN 643
Db 414 -----EHADGSVTQSFDA--VGRLRACTDAAGRTTEYSPDVVTGLITRIT---TPDGR 462
QY 644 TLTYDYELNLDNDRPPPVITTTDVNGQLARNEFDGAGRHSQCCLKSD-----GDGKF 698
Db 463 ASAFYNNHNO-----LTSATGPDGLELRREYDELGRLLIQTETAPDGDITRYRYDNPH 514
QY 699 YTIHTQQVDEQGRHHTSTYSYDLTNGRQOTDPDKVHLSMSKSYDNMGQIANTHWSYGVSE 758
Db 515 SDLPATEDATQSRKTWWSRY--GQLLSFTDCSGYVTRYDHDREFQMTAVHRESGLSQ 571
QY 759 KITVDP1-TLTATKQLQSNNSNVQKVEVTTYTPSQOPIQITLFDGAGHLQ-----SC 810
Db 572 YRAYDSRGOLIAVKDTQGHETRYE-----YNIAGDLTAVIAPDGSR 612
QY 811 HTLTRDGDWRVRKETDAIQCTIYQYDNYNRVLIQITLPGTIVNRKYAPFSTDTLLTDIR 870
Db 613 NGTQYDAMGKAVRTTQG--GLTRSMEYDAAGRVIRLTSENGS-----HTTFRYDVLDRLLIQ 666
QY 871 VNGISLQOQTF--DGLSRLTQSQDGRVWAYTYSAGNDQCPSTVITPDDQFTHYQOPEL 928
Db 667 ETGFGPQRTORYHHDLTGKLIIRSEDEGLV-----THWHYD-EA 702
QY 929 DDVILQVASNEITQOFSYNNPVTGAL--LKAVAEGOSLTPIYY---PSGRL--KMNINDM 981
Db 703 DRLTHTRVKGETAERAWQYDE-RGWLTDISHISEGRHVA--VHYRYDEKGRLTGERTVHPH 760
QY 982 KMSYLV-----TLRLENGYDLDLTOTIQKISRDTGRTVQIKDSSIKITLNYDDLN 1033
Db 761 QTEALLWQHETRHAHYNAOGLAN--RCIPDLSLPAVEWLTLYG-----SGYLAKMGLGD-- 809
```

QY 1034 RHIGSQVTDLATGHLMTTTFVEF--DGLNREIGRKLCDSSGHTLDIQOQSLKTKTQQLANRIV 1091  
Db 810 -----TPLVEYTRDRLHRETLR-----SFGRYELTATAPAGQLOSQ-- 846  
QY 1092 KLVGVLRQTEQYSYSRNRLNQKDCGABCPDKYGHISVITQNFYVDIYGNITACHTTFA 1151  
Db 847 HLNSLLS--DRDYTWNDNGELIR-----ISSPRQ-----TRSYSTTGRLTGVHTTAA 893  
QY 1152 DGTEDHATKFEANPTDPC--QLTEVHHTHPD-----MPDN-----IRLKYDKAGRVIN 1197  
Db 894 -----NLDIRIPIYATDPAGNRLPD--PELHPDSTLSMWPONRIARDAHYLYRDRHGRLTE 947  
QY 1198 ITD-----NHGNTENFTYDTLGRLONGQ-----GSVYGYDPLNRLVSKQKDT 1239  
Db 948 KTDLPPEGVIRTDDETHRYHDSQHRLVHYRTQYEEPLVESRYLYDPLGRVAKRVWR 1007  
QY 1240 LCELE-----YY-----RETMLVNE-----VNGEMIRLLR-----TGETY 1270  
Db 1008 RERDLTGWMSLRKPKQVWYMGDGLRTTIQNDRTRIQIYQPGSFTPLIRVETATGELA 1067  
QY 1271 IAQORA-----SKVLLTGTDSQOSVIL-----TS 1294  
Db 1068 KTORSLADALQOQSGEDGGSVFPVVLQMDLRLESEILADRVSESRRLWASCLTVE 1127  
QY 1295 DRQN-----LSQEA-----YSAYGKHKSTAND---AS 1318  
Db 1128 QMNCQMDPVYTPARKIHLVCHDRGLPLALISKEGTTEWCAYDEWGNLNEENPHQLOQ 1187  
QY 1319 ILGYNGERADPVGVTHLNGVRSYDPTLMRPHTPDSLSPFG--AGGINPYSYCLGDPINR 1377  
Db 1188 LRLPQOQYDEBSGLY--NRHRYDPLQGRVITQD---PIGLKGNFYQYPL--NPVTN 1241  
QY 1378 SDPSG 1382  
Db 1242 TDPLG 1246

## RESULT 4

US-09-252-991A-28424  
; Sequence 28424, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28424  
; LENGTH: 998  
; TYPE: PR1  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-28424

Query Match 3.6%; Score 317; DB 4; Length 998;  
Best Local Similarity 22.0%; Pred. No. 3.1e-15;  
Matches 222; Conservative 129; Mismatches 362; Indels 296; Gaps 47;  
QY 504 EKFIQYRSLIGSQSHVILKIBERHYSATQLNLSLTFQWNTDKSELGRLKQTECTKEN 563  
Db 112 ERWYSYNALG-----LIERADGPRTDVQDVTLYAYD-----SRGNL 149  
QY 564 GHTYSVVKHFTYTKDDTLQQSHSITTHDNFTIHSQVRSYTG--RLFSDDTDKDIVTQ 621  
Db 150 TQVTNALGVTLRGYDERGKPSITDANGVTSSIA-----YGVVDGWLASVSTAGSTTR 204  
QY 622 MSYDKLGRLLTRTLNSGTPYANTLTVDYELANLQDNNRPPFVITTTVDVNGNQLRNEFDGA 681

Db 205 FDYDAVGQITRVRGSG--SWLSYEV-----DDARR--LVAIGNNLGERLEYDVDTK 252  
QY 682 GRHVSOCLKSDGCKGKFTIHTQO--YDEQGR-----HHTSYYS--DYLTNGRQOTDP 730  
Db 253 GNRTAQR1KIDAGS-----LVRQOQWAYDELGRLLRAVAGGQOTRSFAYDLNNDNPVGETNP 308  
QY 731 DKVHLMSKSYDNWGOIANTHWSYGVSEK1TVDPITLTATKQLQSNNSNVOTGKVVTVT 790  
Db 309 RQ--FAHSAQAFDALDRLVGQSDPLGKTRLAYD-----AQNLETKDPRGVTRY 357  
QY 791 PSQQPIQITLDEAGHL-----QSCHTLTRDGDWRVYRKETDAIGCCTIYOYDNNRYI 843  
Db 358 E-----YDGLGNLRLVSPDSGTTTFEHDAGNVIKRTDARGAVTEYRYDALNRLV 408  
QY 844 QITLPDGTIVNRKAPFSTDTLITRVNGISLGOOTFDGLSRLTQSDGGRVWAYTISA 903  
Db 409 ERRSP-----SDPSLDVQYRYDLTADGNQGGIGRLGABGARDS-----446  
QY 904 GNDQCPSTVITPDGQFIHYQYQPELDDAVLOVASNEITQOFSYNPVTGALLKAVAGQSL 963  
Db 447 -----LVRYD--ERGNLVEQVRSIRLQOQLLDRVTTRYDAA-----NQL 485  
QY 964 TPIYVPSGRKVENINDMKMSYLMWTLRGLNGYTDLTGTI1QKTSRDTHGRVTOIK---- 1019  
Db 486 LEIGYPS-----GLAIGY-----PRNAGGQVASVTLAVG 514  
QY 1020 DSSIKTT-----LNYDDLNRHIGSOVTDLATGHLMTTTFVEFDG-----LNREIGHKLCD 1069  
Db 515 DKAPSTLVGOIAYLFPFGPLQ-----LTWNGGITLSREYDQDYQLLRQKV-----560  
QY 1070 SGHTLDIQOQSLKTKQQLANRIVKLVGLQV-----TEQYSYDSRNLNOVK--CDGAEC 1121  
Db 561 -----PWQSDYQ-----HDANGNIQQHRHSLWGLTDIYQDPLDRTEBERGVQGG-- 604  
QY 1122 PTDKYGHSIVTQNFYDIYGNIT--ACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPD 1180  
Db 605 -----RSYADVAVGNRTQSDNPNASGGTASSQDYQVA-----636  
QY 1181 MPDNIRL-----KYDKAGRVINITDNHNTENTFTYDTLGRLOH-----GQGSVYGYD 1227  
Db 637 -PDSNRLTAIGAQAQAVTSDAAG---NLTDRA--ARKLAYDAQGRQLQSLDQOQVAEYRN 691  
QY 1228 PL-NRLVSOKTDTLDCELYRETMLVNRVNGEMIRLRTG-----ETI1AQORA 1276  
Db 692 ALGQRIKVLTPESITTYLYGPDGQLLGEAHDGSGRKLRAQYLYLWLSPLATIDADYDA 751  
QY 1277 S-----KVLITGTDSSQOSVILTSK--QNLQSAYSAYGKHKSTANDAS--ILYNGER 1326  
Db 752 QGKVGNP1LLYLHGDHLDTPLATDASQIAWQSDAFGRGEALSQGSTQVNLRFPGQY 811  
QY 1327 ADPVSGVTHLNGVRSYDPTLMRPHTPDSLSPFG--AGGINPYSYCLGDPINRSDPSG--- 1382  
Db 812 YDAESGLHY--NYFRDYDPTGTRYVESD---PIGLSGGVNTYGVYQGAFLNRLIDPLGLAA 866  
QY 1383 ---HL5WQAWTGIGMGI---AGLLLTATGMAIAAGGIAAIAASTST 1425  
Db 867 IEIDIPKSAVDWIPGNIRLPAGRL-----GGVLLVASISGATPQADSOT 911

## RESULT 5

US-09-328-352-6604  
; Sequence 6604, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6604

```

; LENGTH: 1627
; TYPE: PR1
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6604

Query Match      3.5%; Score 313; DB 4; Length 1627;
Best Local Similarity 20.4%; Pred. No. 1.5e-14;
Matches 303; Conservative 177; Mismatches 506; Indels 498; Gaps 73;

QY 75 IGGIGFNGFGLSVYDRKNLSLSSTGENY--KVIETDKTV-----KLQ 115
DB 369 ITFAMGETPFTHVAQGLGIIQDSISRTVSNLYQMDDAIFGARWVTPPTTKISRFKVT 428
QY 116 OKKLDNLRFEKOLKENCYRIIHKSGDIEVLTFGNNAFPLKVPKLLNPAGHAIY---ID 172
DB 429 SKKQKHQYVNGLEYIC-----LDGRAIDLPLKK-----GOSIYDPIEQ 468
QY 173 WNPPEATQRLNRI-----YDDLGDHDIPLINLEYOGLIKLITLTFPQKGEVYRTEL 224
DB 469 YTYTVLSQHLIAYGEDEKRYEYKYGEDYRLSYIE-----RKNQFKVALR 514
QY 225 F-----LNRLQNSIHNFSGNENPLTWSFGYTPIG-----KNGILQGWITSMTAPGG 271
DB 515 YDHVSTDNKTLISDILFKQDDNLLAHLALQLTFQGLVSDIWTIKNQLDRVLAS----- 568
QY 272 LKETVYNNQGHPPQSANLPVLPLYTLKMQVPGAGQAPQAQAEYSYTSH-----NTV 325
DB 569 -----YDYOQCG-----DLVQATNEBFAASYYYQYTHLITRYTDLT 604
QY 326 GGSNGIWNKLDNLYGLMTEVNGYSTESRRYKDEGHQDQIVRIERTYNNYHLLTSECKQ 385
DB 605 HRGMNLKWGIIPTSKAIEWADNABRASKLEWDK-----NIRKT-----TVLDVEGNS 653
QY 386 QNGYIQTETAYAIIGHFDSQSPQFPLPKTKETWRSADNSYRSEITETTFDSGNPL 445
DB 654 TEHYDIDGTYRIYVPDNFE---ECFFRDDAKNITLHIAKDGSK---TSYTYDERGNVL 707
QY 446 TKVIKDKTKQKIIISPSTHWEYYP-----AGEVNCPPPEY-----GFTRFVKKIIQTPYDSE 498
DB 708 TTTQDDGAT-----SYPEYDEKNQLTGWMDAEQGRWFKQYDQSGNLIKEL----- 752
QY 499 FKXDDPEKFTQYRYSLSGSHVTLKIEERHYSATQLLNSTLFOYNTDKS-----ELGRLLK 554
DB 753 -----DPLK-----HETAYV-----YNAMGLVTSITDAKGGSKSLKYDDQGNLIS 792
QY 555 QTEC-----TK-----GENGKTYSVVH-----KFTYTKODDTLQO----- 584
DB 793 YTDCSGKETKWYDERGRVISIENALNQKVEFYFTELTLENREPIIKGLPLNAFGOLEKI 852
QY 585 SHSITHTDNTIHRSQVRSRYTGRFSDTDTKDIVTOMSYDKLGRLLTFT--LNGSTPYA 642
DB 853 KHADGTEHPF-IHDAE-----GRLLAHVDPKQNTIRYEYDEAGLILSRDALN----- 899
QY 643 NTLTYDYE-----LNNLQDDNPPFVITTTDVNGNQLRNEFPDGNAGRVSOCLKSDGDGKF 698
DB 900 HKLYKWDRLGRLTRLINE-----NGASYQFFYDVASRLV-----KEIDFDGKE 943
QY 699 YTIHTQOYDEQ-GRHHTS--TYSYDLTNGRQOQTD--DKVHLSMSKSYDNWGOIANTHWSY 754
DB 944 TYVH-----YDEKSGOLATSIEVASAYQODLKDRAPKDRIOQFI---FDSMGRLEQRTAGY 997
QY 755 G-----VSEKITVDPI-TLTATPKQIQSNNNVQTGKEVTTYTPPSQOPIQITLFDRA 804
DB 998 GHYGLEEBEQTEEPAYDMGRIIOAKNAQSNLQ-----WFDYDA 1037
QY 805 GHLOSCHTLTRDGDWRVRKETDAIGQCTI--VOYDNYNRVIOITLPGDITVNRKYPAPFT 862
DB 1038 GNLVQEH-----QODYKINKTAVMKHQYDEINRIKTRTPDQGVID----- 1078
QY 863 DTLITDIRVNGISLQOTFDGLSRLTQSDGGRVWAYTYSAGNDQCPSTVITPDGOFIHY 922
DB 1079 -----W-LTYGSHVQ-----SLVINGQDFVSF 1100

```

```

QY 923 QYQPELDDAVLQVASNEITQOFSYNPVTGALLKAVAGOSLTPIIYYPSPGRLXWENINDMK 982
DB 1101 -----BRDDLHREIA-----RHYANGVSOBQOYDLAQLRKSQMM----- 1134
QY 983 KMSYLWTLRGLNGYTD-----LTGTIQKISR-----DTHGRVTOIKDSSI-KTTLNY 1029
DB 1135 -----LSEHENGQYKRNNALEQTSOLVORLYQYDKTGTGELTAIRDTTRGNATYKY 1187
QY 1030 DDLNRHIGSQVTDLATGHMLTTTVBFDGLNRIRGRKLCDSSTGHTLIDIOQSWLKTOQLA-- 1087
DB 1188 DPVGRILLE-----ASSKLGKETFSFPASNIL-----DSVHSQKVQSHSQKLDSTSYG 1235
QY 1088 -NRIYKLANCVLOR--TEQYSYDSNRRLNQYKCGAECPTDKYGHISVITONFYDIYGNIT 1144
DB 1236 YNRLV--NNVVKBYLDQOYDAYQOLTRQKTSQGDG-----NLEWDVYGRMV 1281
QY 1145 ACHTTTFADGTEGHATFKFANPTDPCQLTEVHTHPDMPDNIRLKYDKAGRVINITDNHGN 1204
DB 1282 KSENS-----QYTAEY-----RYDALGRRIQKWSKHH 1309
QY 1205 TENFTYDTLRQLONGGSGVGYDPLNRLVSQKTDLTDCBELYYRETMVNEVRNGEM--IR 1262
DB 1310 T-----GQEQN--IYIGWDG-DTLAYESTEEELTKHYIEKDSFVPMLOAVYLSPIE 1357
QY 1263 LARTGETI-----IAQORASKVLLTGTDSQOQSVILTSQKONLSOE-----AYS 1305
DB 1358 LHQTPDWSDRPYNIIHRDPLMKTEKEGEPDDVMFYCHDLGTPQEMTDHTGAIWKAEYK 1417
QY 1306 AVGKHKSTANDASI-----LYNGERADPVSGVTHLGNVRSYDPTLMRPHTPDS 1355
DB 1418 AMGECKAEEAKGNFENSEIINNRIFQOYFDEETGLHY--NRTRYSPYVGRFVSKD-- 1474
QY 1356 LSPFG-AGGINPYSYCLGDPINRSDPSGHSLSQAWTGMGIAG 1398
DB 1475 --PIGLGNGNVVYA-KNPITWIDSKGLCSITLNRNLG-GVKG 1514

```

# RESULT 6

```

US-09-543-681A-5329
; Sequence 5329, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABI
; FILE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5329
; LENGTH: 1586
; TYPE: PR1
; ORGANISM: Proteus mirabilis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (1576)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-543-681A-5329

```

```

Query Match      3.4%; Score 306; DB 4; Length 1586;
Best Local Similarity 22.1%; Pred. No. 4.9e-14;
Matches 243; Conservative 129; Mismatches 385; Indels 342; Gaps 59;

QY 499 FKDDPEKF-----IQYRS-----LIGSQSHVTLKIEERY-SATQLLNS- 537
DB 486 FEAPHKINTLRKMSADRHNNALHYRYANDGELVQIHDDAYLTDIRLHYDEITORLSV 545
QY 538 -----TLFYQNTDKSELGRLLKQTECKTGKGTYSVVHKKFTYTKQDDTL---QQS 585
DB 546 TRHQGEKTLVITYDAQQ--RLVQVNTADK-----RVTRRFQWDDSEGLMAHMQYA 596

```





Best Local Similarity 18.1%; Pred. No. 2e-12;  
Matches 202; Conservative 179; Mismatches 387; Indels 349; Gaps 49;

QY 440 ESNPLTKVKKTKKTKKISPSHWEYPPAGEVNDPCPPVGFTRFKVKLIQTDPDSEF 499  
Db 337 QSGKISISYGAERVO-----HADFLP-----KIGFS-FIRQ-----YNSQM 373  
QY 500 KDDPEKFIQYRSL-----IGSQSHVTLKIEERHYSATOLLNSTLFOYNTDKSELGRL 552  
Db 374 DEFDSWGCARMWPFNSWNIQNAQGYLFIDSKGRKH---QLPVSIIETVEVPYE-GWI 429  
QY 553 LK-----QTECTGNGKTYSVVHKFTYTKDDTLQOSSHISITH----- 591  
Db 430 IKPLKNGELILDGGEWRSHFQSGKYNLYLVKMMNETSBEILLE-YLLLDHIAYLKV 488  
QY 592 DNFTTHRSQVRSRYTRGLFSDTDKDIVTQMSYDKLGRLLTFLNSGPPYANTLYDYEL 651  
Db 489 INFLLKQAEYELKF-----AFNEQVKIIIAVFL----- 515  
QY 652 NNLODNRPPFVITTTDVNGNQLRNEFDGAGRHSQCLKDSGDGKFYTIHQOYDEQGR 711  
Db 516 ----DDKAEPLA-----RYEYDTQG---NLKAIQDNG-----HTRTYEQNF 551  
QY 712 HHTSYSDYLTNGROOTDPDKVHLSMSKSYDNWQIANTHWSYGVSEKITVDPIFLATK 771  
Db 552 HOLTRYTD--RTGRQNIRESYSTEAKAIEWADDGSPH-----TKLKWHP-----RLR 599  
QY 772 QLOSNNSVQTKVETTYTPSQOPIQITLFDAGHLQSCHT-----LTRGWDVRKE 824  
Db 600 QVA-----YDAYDPTY-----YFDDLDGFTYRTRLDAGRESWYRDGKKRI 645  
QY 825 TDAIGCTIYQYDNNRVITQITLPGTIVNRKYAPFSTDTLITDIRVNGISLGQQTDFGL 884  
Db 646 IDFDRETOEQENDQOLVKIVQPNGGIIRFAY-----NKQ 681  
QY 885 SELTQSD--GGRWAYTYSAGNDQCFSTVITPDGQFIHYQYQ-----PELDDA---VLQV 935  
Db 682 GNLVEIKDPEGSIKWREYDENRN--VSKEINPLGHITQYKYNNDQLVEVIDAKGVKKI 739  
QY 936 ASNEITQFSPNVPVTCALLKVAEQSLTPYIPSGRLKMNINDMKMSYLTWLRGLEN 995  
Db 740 QYNELGQMISYTD-----CSKSSWTEYDEGALTAEQTANNVQYFYSYTKGRDK 790  
QY 996 GYT-----DLTGIOKISRDTHGRVTQIKDSSIKTLLNVDLLNRH----- 1035  
Db 791 GOLQSIILYPDGLKEVFEHDEERLKL-HDTTKGLVTEYKYNQVGLLEQRIIDANRHSVAYQ 849  
QY 1036 -----IGSQVTDLATG-----MLTTTVEFDGLNRBEIGRKICDSSGHTLIDIQSWM 1081  
Db 850 WDKQGRIOKLIQNAQAEYLFYGNPYGLIREQAFDGEKHYG---YNENGRFLFQIRRPNI 906  
QY 1082 KTO-----QLANR-IVKLVGLVORTQYSYDSNRNLNOYKCDGAECPDVKYGHSI--- 1130  
Db 907 LTQFDYADGQIASKSFTHLTGQKTEQFDYNLSQLSRASNEYSQ--IDLYRNALQOL 964  
QY 1131 -----VTQNFYDIYGN-----TACHTTADG-----TSDH 1157  
Db 965 VREHQHYKIPKLPLTAVLHYEYDELGNLIKIRPDGHTLNLHVYSGHIIAIGLNNQEV 1024  
QY 1158 APT-----KFAPTDPDPCQTEVHTHTHPMDPNIR 1186  
Db 1025 VSFQDLDLHRETRLLANGLMOTKVNDVGLLSSQFNQPEQETQYQYQAH-----K 1078  
QY 1187 LKXDKAGRVINITDNHNTENTYDTLGRLONGQ-----SVGYDPLANLVSQKTDLLDC 1242  
Db 1079 YHYDKNYSQVEDSRGLKNTQYDPIGRLIAQSLHKTESFNPAGNLIDSESVLSPA 1138  
QY 1243 ELYYRETMVNRNGEMIRLRTGETIIAQORASKVLLITGDSQSVTLTSDKONLSOE 1302  
Db 1139 QI---KNNLIKSYKGKHQYDVGQNVTEIIQ--AGKULKLTWQNRNRLSRDNGLVTEY 1193  
QY 1303 AYSAYGK--HKSTANDASTLGYNGERA-----DPVSGVTHLNGYRS 1342

Db 1194 GYDVGFRRLYKKTAKELTLFGWDGDLMTWESFKSAQNTYTKHYIYEPDSFVPLLOQGYKD 1253  
QY 1343 YDPTLMRPHHTDLSLSPFGAGGINPVSYCLGDPPI-NRS 1378  
Db 1254 F---IQLIETPD-----YQEYQTKFYS-IYKDPVWNRN 1282

## RESULT 10

US-09-252-991A-23640  
; Sequence 23640, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23640  
; LENGTH: 974  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23640

Query Match 3.1%; Score 277.5; DB 4; Length 974;

Best Local Similarity 20.4%; Pred. No. 3.5e-12;

Matches 218; Conservative 141; Mismatches 377; Indels 335; Gaps 48;

QY 504 EKFIQYRSLGSSQSHVTLKIEERHYSATOLLNSTLFOYNTDKSELGRLKQTECKGEN 563  
Db 112 ERWYSYNALG-----LIKRADGPRTDQDVLTLYAYD-----SRGNL 149  
QY 564 GKTSVSVHKFTYTKQDDTLQOSSHITTHDNFTIHRSQVRSRYTG--RLFSDTDTKDVI 621  
Db 150 TQVTNALQOVTRLGDYDERGKPGSITDANGVTSSLA---YTGVDGWLASVSTAGSTR 204  
QY 622 MSYDKLGRLLTRTNSGTPPYANTLYDYELNNLQDNRPPFVITTTDVNGNQLRNEFDGA 681  
Db 205 FDYDAVGQITRVTGRDG---SWLSYEV-----DDARR--LVAIGNNLGERLEYDVT 252  
QY 682 GRHYSQCLKSDGDKGFYTIHQO--YDEQGR-----HHTSYVS--DYLNGRQQTDP 730  
Db 253 GNRTAQRICKASGS---LVRCQQWAYDELGRLLRAVGAGGQTRSFAYDLNDNPFVETNP 308  
QY 731 DKVHLSMSKSYDNWQIANTHWSYGVSEKITVDPIITLTATKLOQSNNSNVQTKVETTYT 790  
Db 309 RQ--FAHSQAFDALDRLVGQSDPLGGKTRLAYD-----AQDNLTEVKDPRGVTT 357  
QY 791 PSQOPIQITLDEAGHL-----QSCHTLFRDGHWRVRETDAIGCTIYQYDNNRV 843  
Db 358 E-----YDGLNLTSLVSPDSGTTTFEHDAGNVIRRTDARGAVTEYRDALNRLV 408  
QY 844 QITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTDFGLSRLTQSDGGRVWATYSA 903  
Db 409 ERRSPSDSLDVQYR-----YDLTADGNK-----GIGRL-GAIDGAR----- 444  
QY 904 GNDQCFSTVITPDGQFIHYQYQPELDDAVLQVANSNEITQQFSYNFVTGALLKVAEGOSL 963  
Db 445 -----DSLVYRYD-ERGNLVEQVRSIRLQOQLDRLVTRYDAA---NQL 485  
QY 964 TPIYPSGRLLKMNINDMKMSYLTWLRGLENYTDLTGTQKISRDTGTRVTOIKDSSI 1023  
Db 486 LEIGYPS-----GLAIGY-----PRNAGGQVASV----- 509  
QY 1024 KTLNLYDDLNRHIGSQVTDLATGHTMLTTTVEFDGLNRBEIGRKICDSSGHTLIDIQS 1083  
Db 510 --TLAVGDKAPSTLVGQIAYLPFGPLRLUT-----WNGITLSREYDQD 551



QY 1084 QOLANRTV-----KLVGLQR-----TEQSYDSRNFLNOYK-CDGAECPDCKY 1126  
DB 552 YQLLRQKQWQSDYQHDANGNIQOHRHSLWGLDYOYDPLDRLLTEERGVOGG-----604  
QY 1127 GHSIVTQNFYDIYGNIT-ACHTTFADGTDHATFKFANPTDPCQLTEVHHTHPDMPDNI 1185  
DB 605 -----RSYAYDAVGNRTQSDNPASGCTASSQDYQYA-----PDSN 640  
QY 1186 RL-----KYDKAGRVINITDNHGNENTFTYDTLGRLOQ-----GQGSVGYDPL-NR 1231  
DB 641 RLTAIGAQAQVTSDAAG---NLTDQRA-ARKLAYDAQGRQSVSLDQGVVAERYNALGOR 696  
QY 1232 LVNSOKTDLTDLCELYREYTMVNRNGEMIRLRTG-----ETIIAQORAS-----1277  
DB 697 IVKLTPSEVTTYLYGPDGQLLGAEGHGGGRKLRAQYLLWLSPLATIDADYDAQKVG 756  
QY 1278 --KVLITGTDSQSVILTSK--QNLQSQAYSAYGKHKSANDAS--ILYNGERADPVS 1331  
DB 757 NPFLLYLHGHDLTPLRATDASQIAWQSQDAFGGEALSGSTQVNLRFPPQYYDAES 816  
QY 1332 GVTHLGNVRSYDP-----TLMRFTPD-----SLSPFGAGGINP--1366  
DB 817 GLRY--NYPRDYDPETGRYVESDPVETLKLNNPENTFLNSGSMQLQATPYWEHGFTPNH 874  
QY 1367 -YSYCLGDPINRSD-----PSGHLW-----QAWTGIGMGIAGLLTIAT 1405  
DB 875 NYTYSNNTAKSDRGLSPNPTDNLITYPTDNTCTCLECKKTKGKSF-----925  
QY 1406 GGMAIAAGGIAAIASTSTALAFGALSVTSDITSIVSGALEDASPKASS 1456  
DB 926 -----LVGALCSKATTPF-FGVVCNNTIVNICGASCQECNRAPS 965

## RESULT 11

US-09-543-681A-4476  
; Sequence 4476, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709,1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4476  
; LENGTH: 1665  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-4476

Query Match 3.1%; Score 271.5; DB 4; Length 1665;  
Best Local Similarity 20.7%; Pred. No. 2.6e-11;  
Matches 233; Conservative 148; Mismatches 350; Indels 395; Gaps 66;  
QY 508 QVRYSLIGSQSVHTLXIBRRHYSATOLLNSTLFPQYNTDKSELGLLKQT-----EC 558  
DB 608 QHRIDFIYEQSQLSV-KRHDGIALIL-----EYQDK--LVRICSESTTSRPFVRC 659  
QY 559 TKGENGKTVSVVHKFTYTKQDDTLQOQSHSITTHDNFTIHRSQVRSRYTGR-----LPSDTD 614  
DB 660 EYDTHG-----YLSQCHAYQNH-----LWHYSPGELMVAGWDDTD 695  
QY 615 TKDIVTQMSYDKLGRLLTRTLNSG-----TPYANTLTY-----DYELNNLQDDNRPP 661  
DB 696 STELT--IDYDEQGRVATHSPSGFWNDRFIYDDYQRMWTYIDAEGGFSRYYNDDN---750  
QY 662 FVITTT-----DVGNGLRNEFDGAGRHSVQCLKD-----SDGCKFY 699  
DB 751 -LVTRTIDPLWRETYTEWEQORKIABINEIGERTEY---GYHNGLLAVIYLPDGGKAIY 806

## RESULT 12

US-08-851-567B-61  
; Sequence 61, Application US/08851567B  
; Patent No. 6528484  
; GENERAL INFORMATION:  
; APPLICANT: Ensign, Jerald C  
; APPLICANT: Bowen, David J  
; APPLICANT: Petelli, James  
; APPLICANT: Fatic, Raymond  
; APPLICANT: Schoonover, Sue  
; APPLICANT: firenche-Constant, Richard  
; APPLICANT: Rocheleau, Thomas A.

APPLICANT: Blackburn, Michael B.  
APPLICANT: Hey, Timothy D.  
APPLICANT: Merlo, Donald J.  
APPLICANT: Orr, Gregory L.  
APPLICANT: Roberts, Jean L.  
APPLICANT: Strickland, James A.  
APPLICANT: Guo, Lining  
APPLICANT: Ciche, Todd A.  
APPLICANT: Sukhapinda, Kittisri  
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: IN  
COUNTRY: US  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,567B  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 514  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/063,615  
FILING DATE: 18-MAY-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/395,497  
FILING DATE: 28-FEB-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,255  
FILING DATE: 06-NOV-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,423  
FILING DATE: 28-FEB-1996  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/705,484  
FILING DATE: 28-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27396  
REFERENCE/DOCKET NUMBER: 960296.93804  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1043 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-851-567B-61

Query Match 2.7%; Score 238.5; DB 4; Length 1043;  
Best Local Similarity 20.4%; Pred. No. 4.2e-09;  
Matches 221; Conservative 130; Mismatches 378; Indels 353; Gaps 51;

QY 612 DTDTKDIVTOMSYDKLGRLLTRTLNSGTPYANTLTVD---YELNNLQDDNRPFPVITTTD 668  
DB 38 DTDTR--VTRHYDARGHL-----NYSIDPRLYDAKQADNSVRFV-WQHD 81  
QY 669 VNGNQLRNEFDGAGRHSVQCLKSDSG-----DGKFTYHTTQY 706  
DB 82 LAGHALRTESVDAGRTVA--LNDIEGRSVMTNATGVQTRRYEGNTLPGRLLSVSEQVF 139  
QY 707 DQGRH-----HTSYSDYLNGR---QQTDDPKVHLSKSY----- 741  
DB 140 NQESAKVTRFTWAGNTTSEKYNLSGLCIRHYDTAGVTRLMSSQSLAGMLSQSHQLAE 199

QY 742 ---DNWQIANTHSYGVSEKITVPITLTATKQLQSNNSNVQTKGVYTYTPSOQPIQI 798  
DB 200 QGEANWSGDDETVWQMLASEVYT-----TQSTTNAIGA---LLTQTDKGNQIR 246  
QY 799 TLFDEAGHLQSCHTLTRGDWRVRKETD-----AIGCCTIYQDYNVRVQITLIP 848  
DB 247 LAYDIAGQL-----KGSWLTVRGQSQBVIVKSLWSAAGHKLREHGN----- 289  
QY 849 DGTIVNRKYAPFSTDTLTITDIRVNGISLGQQTFDGLSRLTOSQDGRVWAYTYSAGNQC 908  
DB 290 -GVVTEYSYEP-ETORLLI-----GIT-----TRRABGSQSGARVL----- 322  
QY 909 PSTVITPDGQFIHYQYQPELDDAVLQVASNEITQFSYNPVTGALLKAVAEGQ-----SLT 964  
DB 323 -----QDLRYKYDPV--GNVISIHDAEATFRWENQ-----KVEPENRYVDSLY 365  
QY 965 PIYPSGRLKMNINDMKOM--SYLWTLRGLENGHTDITGTIOKLSRTHGRVTKOSS 1022  
DB 366 QLMSATGR-EMANIGQSNQLPSPVIVPTDDSTYTN---LRTYVDRGGLVQIRHSS 421  
QY 1023 IKTTLNY--DDLNRHIGSOVTDLATGHMLTTTVEPDLNREIGRKLCDSSGHTLDTQQSWL 1081  
DB 422 PATQNSYTTDIT--VSSRSNRRAVLSTLTDTPTRVDA-----LFDSGGH-----QKMLI 467  
QY 1082 KTQQLANRIVKLVGLQTEQYSYD--SRNRLNQYKCDG---AECPTDKYGHISVITQNT 1136  
DB 468 PQQLNDWNI---RGLQRVTPVSRENSDSEWRYSSDGMRLLVKVSQEQTKNSTQVRVT 524  
QY 1137 YDIYGNITACHTTFADGTEDHATKFNPTPCQLTEVHHTHPDMPNIRUKYDKAGRVI 1196  
DB 525 Y-----LPGLELRTTGVAKTTEDLQVITVGEAG---RAQVRVLHWESGKPT 568  
QY 1197 NITDNHGNENTFYDTLGRQLNGQGSVGYDPLNRLVSKQKTDLDCELYRYETMLVNEVR 1256  
DB 569 DLDNNQVR-----YSYD--NLGSSQLE----- 589  
QY 1257 NGEMIRLLRGTETIIAQORASKVLLITGTDSSQSVILTSDKQNLISOEAYSAYG-----KH 1310  
DB 590 -----LDSEGOILSQEYYPYGGTAIWAAR 614  
QY 1311 KSTANDASILYNGERADPVSGVTHLNGYRSYDPTLMRFTTPDSLSRPG-AGGINPYSY 1369  
DB 615 NQTEASYKFIYSGKERD-ATGLYYY--GYRYQPVWGRWLSAD---PAGTVDDLNLRYM 668  
QY 1370 CLGDPINRSDPSG-----HLSWQAWTGIGMGIAGLLLTATG---GMAIA---AAG 1414  
DB 669 VRNPITLTDHDLAPSPNRRNTFWFASFLFRPDEGMSMRGQKIGRAIAGIAG 728  
QY 1415 GIAAAIASTTAL--AFGALSVTSDITSIVSGALEDASPKASSILGWVSMGM----- 1465  
DB 729 GLAATIAATAGNAIPVILGVAAGVAGIGALMGYNVGSLEKGGALLARLVQKSTLVQSA 788  
QY 1466 --GAAGLAESAIGK----GTKLATHLGFABEDGENALLKSTSESSRIKWKVTRSLDREIV 1519  
DB 789 AGAAAGASSAAAYGARAQGVGVASAAGAVT-----GAVGSWINNAD 829  
QY 1520 RNESQVVKDHSRGVYTNFMKGKGEQAILVH-----GDKDGLYHTTGK-----HNGKG 1568  
DB 830 RGIGGAIGAGSAGVTIDTLMGTA--STLTHEVGAAGGAAGGMITGTQGSTRAIGHAGIG 887  
QY 1569 PY 1570  
DB 888 TY 889

RESULT 13  
US-09-328-352-5545  
; Sequence 5545, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS



Db 697 CEVRYTYDEGNL-RLESQVHADGWVYSVEHSHDALGVQTSRYGD-----APPAVAVLT 750  
QY 1044 ATGHM---LTTTVEF---DGLNREIGRKLCDSSGHTLIDIQSWLTKQ-----LAN 1088  
Db 751 GPGLHGLVGA VELAFAERDALHREVRDA-----RRDQDDALFTQERQHAPLRLQR 804  
QY 1089 RIVKLVGLVQRTQSYSDSRNL-----NOYKCDGAECPTDKYGHSI-----V 1131  
Db 805 SRLLAGGPDWQGYRYDGLQLVGIDDDNQ-----PSVRYEYDLGRLLASRAGAA 857  
QY 1132 TQNTFYDIYGNITACHTTFADGTEDHATFKFANPTDPCQLTEVHTHPDMPONIRLKYDK 1191  
Db 858 ASTYRYDAAGN-----RLEGVGEYAR-----EDARQAFENELRYSGFSKSEVRANQ 904  
QY 1192 AG-----RVINITDHNGTENFTYDTLGR-----QNGGSGYVDPLNRLVSQKT 1237  
Db 905 AGEPPARWAGNRVERIAGN-----RYRFDALGNLVERIGADGERLRLAYDGAQRLVHLTR 959  
QY 1238 DTLD-----CELYRETMLVNEV-----RNG-EMIRLLRTGETIIAQORASKVLLTGTDSQ 1287  
Db 960 DYADGTRLEARYYDALSRRIAKVLRDGVQVRFQWGDQCAEAFARELRTTHERPG 1019  
QY 1288 QSVIL-----TSKQNLQSO--EAYSAYGKHKSTANDASILGYNGERADPVSGVTHLGN 1338  
Db 1020 GFVPLRLRQACEPDPPELLQLRQAPAEQGPL-----PAQCVPALGE 1062  
QY 1339 GYRSYDPTLMRHTPDLSPFPGAGGINPYSYCLGDPINRSDPSGHLNWAQ-----WTGI 1392  
Db 1063 ARIAF-----FHTDH-----LGTPLQSLDSBERGQLRWQGVPPDRAV 1098

RESULT 15  
US-09-328-352-5486  
; Sequence 5486, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5486  
; LENGTH: 3290  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5486

Query Match 2.5%; Score 222.5; DB 4; Length 3290;  
Best Local Similarity 18.8%; Pred. No. 5.4e-07;  
Matches 370; Conservative 249; Mismatches 656; Indels 697; Gaps 101;

QY 12 ITMSDNEF-----FTQANNFTSAVSG-----GVDPR 39  
Db 459 LTLSDLNDTVGSATDQITODKNFLKLEQETGSGQVTLVSTDEGKTWQETTVAQKDLAD 518  
QY 40 GLYNTIQLTHGVNGN-----LGPTLPLT--LSYSPLNKTDIGFGIFNGLSVYD 89  
Db 519 GIYQKAVVDVAGNISETAIOKVVDVDTTAPQAGKLTLSLNDT-----GVSATD 568  
QY 90 ----RKNS-----LLSLSTGE-----NYKVIETDKTVKLQKKDLNLRPEKOLKENCYRII 136  
Db 569 QITQDSNFTKLQVPIVIGEQAALLDHYEVSDEGKTWQETTAD-----QKLDADIYQ-- 622  
QY 137 HKSQIEVLTGNNNAFDLVKPKLLNPAGHAIYIDMNFEPATQPLNRYDLDLGHDIPL 196  
Db 623 YKA-----VVTDLAGNISESAIOKVVDNS-----LNVESTTVIVKPTED----- 663  
QY 197 LNLVYQGLIKTILTLTPQCKEGRYRFLRNLQANSIHFNLSGNENP-----LTWMSFGYTP 252  
Db 664 -----NTLSVEKQDVISIRLEIANLPTDLNS-----SLTSVNTTLENVTYNFHFDE 710

QY 253 IGKNGILGOWIT-----SMTAPGCLKETVYNSNNQGHHPQSAN 292  
Db 711 VTQ-----EWTETPABFLMSVEPQTNISIDISLTDQAGNTAIIHTHQ-----YNDVPTPN 762  
QY 293 LPLVPLVLMKQVPGAGQPAIQASYSYTSYHNVGSGNGIWNKLDNLYGLMTEYNYGST 352  
Db 763 SPTLDSLT- FNNIDGA-----IISGAYKSKVDIYNKNGDWLASIT- 804  
QY 333 ESRRYKDEKH-----DQIVRIERTYNNYHLLTSECKQONGYI--QTTETAYAI- 400  
Db 805 -----NEEGKFTLQDLSINTNQEVYAVATNGY-----SSENSIGLVEVEPAISIT 851  
QY 401 -----IGHNPDQSPQPOLPKTKTETWRSADNSYRSIETTTTDESG----- 442  
Db 852 RISPEGVISGATEGSHFIV-----KQONGNILQEFNSNVFDSGGITPFSVMALGEV 903  
QY 443 -----NPLTKVIKDKTKIISP-----STHWEY-----YPPA-----GEVD 474  
Db 904 RPFILSLDQPL-----BEGAQIIISTDKDNISGHPQYITADYTPAVFLETQFDISGETL 958  
QY 475 NC-PPPEYGFTRFVK---KIOTPYDSEFKDDPEKFIQYR-YSLIGSOSHVTLKIEERHY 529  
Db 959 SVHVNEPNSFIRAFSGEGNLIATGF-----TDEQGFASLQVQFLKEGETVTVQVVDKNQ 1013  
QY 530 SATOLL-----NSTLFPQYNTDKSELGRLLKQTECTKGE 562  
Db 1014 NISETLIIEVPNFAYIPHVERITOEGLSGVAEDNSTVIVRADAGNELKV----- 1063  
QY 563 NGKTVSVVHKFTYTKQDDTLQOQSHSIITHDNFTIHRSQVRSRYTG-----R 608  
Db 1064 -----TLGDD-----NSWSDPFSHLSVNRPLIDGEEKISVQIIDNKGLMSPQN 1107  
QY 609 LFSDDTDKDIYQMSYDKLGRLLTRTLNSGTPYANTLYD-----YELNNLODD--- 657  
Db 1108 IIVDLTPPPAPTFLNFNDAGDLV---YGHAEPPFSEILVKDQGGNLFKFWNNWNTDESGS 1164  
QY 658 -----NRPPPVITTDVNGN-----QLRN-----EPDG-----AGRHV 685  
Db 1165 FSIELGFTLWAEYVYVATDVNGVNSLAAQIAPNAPAPYVDSFTSDGVISQAEENS-1224  
QY 686 SQCLKSDGDGKFYTIHTQQYDEQGRHHTSYSDYLTNGRQQTDPDKVHLSMSKSYDNWG 745  
Db 1225 TLWVKDAKD-----VVAEIKVGEDNGWNGSSYFK-LQLDRPLVDGEQFFLSIK---DARG 1276  
QY 746 QIANTHWSYGVSEKITVDPIPLT-ATKQLOQSNNVQTG-KEVTTYTPSQOPIQITLDE 803  
Db 1277 QVS-----ADTVITADTVAPTASNLFVSEDSGYLTGVAELNT-----TIQVFDH 1321  
QY 804 AGHLQSC--HTLTRDGDWRVRKETDAIGQCTIY-----OYDYNVRVIQ 844  
Db 1322 NGQLVNIWNNTINGD-----GTFIYLGNNLHGEAFTVTVKQAGNVSEALS 1369  
QY 845 ITLPDGTIVNRKYAPFSDTLITDIRVNGISLQO-----TFDGLSRUTSQDQGRV 896  
Db 1370 INAPLDDI-----APNPKNILLD--ANGNFTAQAEANSQIEVFDSLQ-----NOTG--- 1415  
QY 897 WAYTYSAGNDQCPSTVITPDQGFTH-YOYQPELDDAVLOVASNE-----ITQOFSYN 947  
Db 1416 WGSTDSAGN-----ASGSFNQTYLHGBELTFVVIDRAGNRSIEBFKQNALIDTTAPN 1466  
QY 948 PV-----TGALLKVAEGOSLTPYVPSGRKLWENINDMKMSYLWTLRGLENGYSYDL 1000  
Db 1467 PIENIIFNENGQSFQAQAEAGSSIDLVDOTG-----NKIGF-----GYTDS 1507  
QY 1001 TGTQKTSRDT--HG-----RVTOIKDSSIKTTLNYDDLNRRHTGSOVTDIAT 1045  
Db 1508 SGNVSGYFQQVYLHGBELTFVVIDRAGNRSAEVQSA-----LNDVAPNPIENILLDL-N 1562  
QY 1046 GHMLT-----TTVEFDGLNRE-IGRKLCDSSGHTLD-IQOSWLKTOOLANRIVKLGV 1097  
Db 1563 QONFTAQAEANSQIEIKNNNGDVVGYGSADSGAGNVGYLYQVHLHGBELTFIVVDRA-- 1620

QY 1098 RTEQSYSDSRNLNQYK-----CDGAECPDKYHSIVTONET----- 1136  
Db 1621 -----NRSTEYKQNALIDDIAPNPIENIVLDINGQNFQAQAEANTQIEVKNVAVG 1669  
QY 1137 -----YDIYGNITA-CHTFADGTEDHATKFKAN-----PTDPQCLTEVHHTHPDMPD 1183  
Db 1670 EIVSGVVDGAGNVSGLYQVYLHGB--LTFVVDVDRAGNRSTEVKQNALIDDIAPNPIE 1727  
QY 1184 NPLKVDKAGRVINITDNHGNTEFT-----YDTLGRQLONGQSGVYQVPLNRLVSOQTDT 1239  
Db 1728 NIL-----DANGQNFQAQAEANTQIEVKNAAAGEVIG-----SGSTDs 1765  
QY 1240 L-DCELYRETMLVNEVRNCEMIRLL---RTGETIIAQOPASKVILTTGDSQOSVILTSD 1295  
Db 1766 MGNVSGYFYQVYL-----HGEELTFVVDVDRAGNRSTEVKQNALIDDIAPNATENIIFNEN 1820  
QY 1296 KQNLQSEA-YSAYKHKSTANDASILGYNGERADPVSG-----VTHLGN 1338  
Db 1821 GQNFQAQAEANSKVEVKNAAGEVVGSGY-VDSAGNVSGYLNQVYLKGBELTFVVIDOAGN 1879  
QY 1339 GYRSYDPTLMRHTPDSLSGPFAGGI-----NPYSYCLGDP-----INRSDPSGHL----- 1384  
Db 1880 --RSIEVKOTAF--LDNTAPENATNLVFSGEDGSLSGMAEPNATIQIFDQYGQLLNQWNN 1935  
QY 1385 --SQOAWTGIGMG---IAGLLLTATGCMIAIAAGGIAAIASTSTTALAFGALSVTSDI 1439  
Db 1936 NVNWDGTFNLYLNSNTMHGEVFKV---VVVDQAGNLSGEVT----- 1973  
QY 1440 TSIVSGALDASPKASSILGWVSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKS 1499  
Db 1974 ---VKAPLDDIAPVAASDLVFNEDGSSLSGVAEP-----NTFIQIFDQNGQQ--MNT 2020  
QY 1500 TSSESRKMGVTRSLDREIVRNEEGQVIKDHRSRGYTDNFMKGEGEQAILVHGDKDGFLYHT 1559  
Db 2021 WQSQVNAADGTFITFFGTYNLHGEFTVI-----VKDLAGNVSEAVSV----- 2062  
QY 1560 EGNKHNGKGPYTRHTPEQLVDYLYKDNINVDLTQGD--KPVHLLSCYKSSG 1609  
Db 2063 -----KAPLDDIAPNPIKNIVFDANGQSFTAQAEANSQIEIFDSFGSQIG 2107

Search completed: February 9, 2005, 18:20:23  
Job time : 74 secs

**THIS PAGE BLANK (USPTO)**